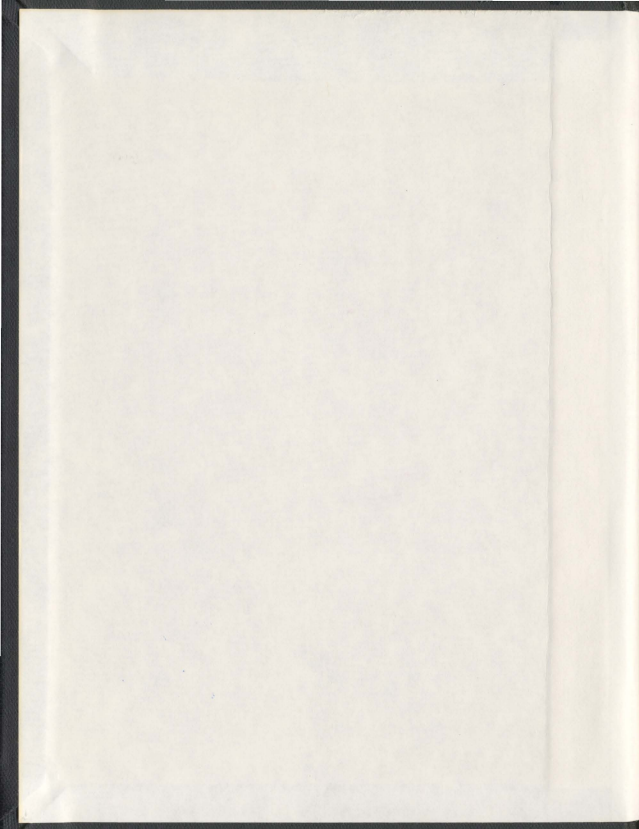


OLIGONUCLEOTIDES SEQUENCING AND GLYCATION
SITES DETERMINATION OF NEOGLYCOCONJUGATE
VACCINES USING TANDEM MASS SPECTROMETRY

FARID JAHOUH



001311



**Oligonucleotides sequencing and glycation sites determination of
neoglycoconjugate vaccines using tandem mass spectrometry**

by

©Farid Jahouh

A Thesis submitted to the

School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

February 2012

Department of Chemistry

Memorial University of Newfoundland

ABSTRACT

Mass spectrometry (MS) has turn out to be a fundamental technology which has produced tremendous strides in the biological sciences. In the last three decades, MS has witnessed convincing growth. Mass spectrometry has rapidly evolved to the forefront of analytical techniques; its ability to analyze proteins, deoxyribonucleic acid (DNA) and carbohydrate biomolecules has been a major driving force in the field of proteomics, genomics and glycomics. The level of performance that is achievable with MS today allows scientists to study various biomolecules in ways that were inconceivable 25 years ago. In this thesis, two different kinds of biomolecules will be covered: oligonucleotides and carbohydrate-protein neoglycoconjugates.

First of all, the differentiation and sequencing of three constitutional isobaric 18-*mer* DNA oligomers, namely: **GATTCATAGCTACGAATC 1**, **AATTCGTAGCTACGAATC 2**, and **AATTCGTACCTACGAATG 3** were investigated using electrospray ionization (ESI) mass spectrometry with a hybrid QqTOF-MS/MS instrument. Our main objective was to differentiate and to sequence these oligonucleotides. The hypothesis was that the single scan ESI-QqTOF-MS analyses of the DNA oligomers will afford the same series of deprotonated molecular ions and that low-energy collision tandem mass spectrometric analyses (CID-MS/MS) will allow us to differentiate and to sequence the three DNA oligomers using the Mongo Oligo Software. Another objective was also to establish a mass fingerprint of these three oligonucleotides. This will aid to characterize the structure of covalently linked carcinogens to these three isobaric oligonucleotides.

The second part of the thesis, which includes several chapters, is concerned with the use of MS for the localization of the glycation sites of neoglycoconjugate vaccine models formed by the covalent attachment of specific immunogenic carbohydrates obtained from different virulent bacteria to the protein carrier bovine serum albumin (BSA). This can be accomplished using a methodology based on squaric acid chemistry which was developed to conjugate carbohydrate antigens to lysine residues in proteins. Three different examples of hapten-BSA glycoconjugate vaccines were used: the *Vibrio cholerae* O1 antigenic specific hapten conjugated to the BSA, the tetrasaccharide side chain of the *Bacillus anthracis* exosporium conjugated to the BSA, and finally the model β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside (β -lactoside) conjugated to the protein carrier BSA. Initially, the recording of the MALDI-TOF-MS of the different hapten-BSA glycoconjugate vaccines allowed us to determine the hapten-to-BSA ratios. The neoglycoconjugates were then digested and analyzed by MALDI-TOF/TOF-MS/MS and LC-ESI-QqTOF-MS/MS for the determination of the glycation sites. Our hypothesis was that the trypsin digestion will not allow a complete digestion of the protein, since the glycated lysines are not reactive with trypsin. Thus we had to use another protease, the GluC V8 which is known to digest proteins at the C-terminus of the aspartic acid and glutamic acid residues. We also suspected that the LC-ESI-QqTOF-MS/MS analysis was going to afford the identification of a higher number of glycation sites than the MALDI-TOF/TOF-MS/MS analyses.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my supervisor, Dr. Joseph H. Banoub, Adjunct Professor of Chemistry at Memorial University of Newfoundland and Head of Special Projects, Science Branch, Department of Fisheries and Oceans Canada, for welcoming me into his lab during these four years, for his advice, and supervising my studies.

I also would like to thank Professors Travis Fridgen and Graham Bodwell (supervisory committee) for reviewing a draft of the thesis and for their advice.

I would like to thank the Department of Fisheries and Oceans (DFO), St John's for providing the resources which enabled me to perform my research studies, and the Chemistry Department of the Memorial University of Newfoundland.

I wish to thank Howard Hodder, Georges Sheppard and Bob Whalen for their assistance in the Department of Fisheries and Oceans.

Finally, I would like to express my gratitude to the School of Graduate Studies (SGS) of the Memorial University of Newfoundland and National Sciences and Engineering Research Council (NSERC) for the financial support throughout my degree program.

To my family.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	xi
LIST OF TABLES	xvii
LIST OF ABBREVIATIONS	xix
CHAPTER 1: Introduction	1
1.1. Biomolecules	2
1.1.1. Oligonucleotides	2
1.1.2. Carbohydrates	4
1.1.3. Bacteria	8
1.2. Mass spectrometry	17
1.2.1. Ion sources	17
1.2.2. Mass analyzers	22
1.2.3. Ions in the gas phase	31
1.3. Mass spectrometry analysis of biomolecules	33
1.3.1. Mass spectrometry analysis of oligonucleotides	33
1.3.2. Mass spectrometry analysis of proteins and peptides	36
1.3.3. Mass spectrometry analysis of glycoconjugates	37
1.4. Co-authorship statement	38
CHAPTER 2: Differentiation and sequencing of three constitutional isobaric 18-mer	
DNA oligomers using low-energy collision tandem mass spectrometry	60
2.1. Introduction	62

2.2. Material and methods.....	67
2.2.1. Sample preparation	67
2.2.2. Electrospray quadrupole orthogonal time-of-flight mass spectrometry	67
2.2.3. Low-energy collision CID-MS/MS analyses.....	68
2.3. Results and discussion	69
2.3.1. ESI-QqTOF-MS analyses	70
2.3.2. Low-energy collision CID-MS/MS analyses.....	72
2.3.3. CID-MS/MS analysis of the $[M - 8H]^{8-}$ deprotonated molecules at m/z 683.9881 extracted from the isobaric 18-mer DNA oligomers 1-3	72
2.3.4. CID-MS/MS analysis of the $[M - 9H]^{9-}$ deprotonated molecules at m/z 607.8775 extracted from the isobaric 18-mer DNA oligomers 1-3	77
2.3.5. Common product ions present in the product ion scans of the oligomers 1-3	79
2.3.5. Common product ions present in the product ion scans of the oligomers 1-3	80
2.4. Conclusion	86
CHAPTER 3: Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of <i>Vibrio cholerae</i> O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry.....	
3.1. Introduction.....	95
3.2. Material and Methods	99
3.2.1. Preparation of the hapten-BSA neoglycoconjugates	99
3.2.2. Digestion	100
3.2.3 MALDI-TOF-MS and MALDI-TOF/TOF-MS/MS analyses	101
3.3. Results.....	101

3.3.1. MALDI-TOF/TOF-MS analysis of BSA and the hapten-BSA neoglycoconjugates.....	101
3.3.2. High-energy CID-MS/MS analyses of the digested BSA and hapten-BSA conjugates	108
3.4. Conclusion	117
CHAPTER 4: Determination of the glycation sites of <i>Bacillus anthracis</i> neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF- tandem mass spectrometry	
4.1. Introduction.....	126
4.2. Material and Methods	129
4.2.1. Preparation of the hapten-BSA neoglycoconjugate.....	129
4.2.2. Digestion	129
4.2.3. MALDI-TOF/TOF-MS and MALDI-TOF/TOF-CID-MS/MS analyses.....	130
4.2.4. LC-ESI-QqTOF-CID-MS/MS analysis	131
4.3. Results.....	132
4.3.1. MALDI-TOF/TOF-MS analysis of BSA and the hapten-BSA neoglycoconjugates.....	133
4.3.2. PMF analysis using MALDI-TOF/TOF-MS of the digested hapten-BSA neoglycoconjugates.....	134
4.3.3. High-energy MALDI-TOF/TOF-CID-MS/MS analyses of the glycated peptides	137
4.3.4. LC-ESI-QqTOF-CID-MS/MS analysis of the digested carbohydrate-BSA conjugate	144

4.4. Conclusion	153
CHAPTER 5: Revealing the glycation sites in neoglycoconjugate models formed by	
conjugation of the antigenic monosaccharide hapten of <i>Vibrio cholerae</i> O1 serotype	
Ogawa with the BSA protein carrier using LC-ESI-QqTOF-MS/MS.....	161
5.1 Introduction.....	163
5.2. Material and Methods	165
5.2.1. Preparation of the hapten-BSA neoglycoconjugate.....	165
5.2.2. Neoglycoconjugates digestion with different proteases	165
5.2.3. LC-ESI-QqTOF-CID-MS/MS analysis	166
5.3. Results.....	167
5.3.1. Mass spectrometry determination of the glycation sites on the	
neoglycoconjugate possessing a hapten:BSA ratio of 4.3:1	168
5.3.2. Mass spectrometry determination of the glycation sites on the	
neoglycoconjugate possessing a hapten:BSA ratio of 6.6:1	185
5.3.3. Mass spectrometry determination of the glycation sites on the	
neoglycoconjugate possessing a hapten:BSA ratio of 13.2:1	186
5.4. Conclusion	191
CHAPTER 6: Determination of glycation sites by tandem mass spectrometry in a	
synthetic lactose-BSA conjugate, a vaccine model prepared by dialkyl squarate chemistry	
.....	195
6.1. Introduction.....	197
6.2. Material and Methods	200
6.2.1. Hapten-BSA glycoconjugate preparation	200

6.2.2. Digestion	200
6.2.3. MALDI-TOF-MS analysis.....	202
6.2.4. LC-ESI-QqTOF-MS/MS analysis	202
6.3. Results.....	203
6.3.1. MALDI-TOF-MS analysis of the hapten-BSA glycoconjugates.....	203
6.3.2. LC-ESI-QqTOF-MS/MS analyses of the tryptic and GluC V8 digests of the neoglycoconjugate with a lactose:BSA ratio of 5.1:1 and 19.0:1	205
6.3.3. Mass spectrometry determination of the glycation sites on the neoglycoconjugate with a hapten:BSA ratio of 5.1:1	206
6.3.4. Mass spectrometry determination of the glycation sites on the neoglycoconjugate with a hapten:BSA ratio of 19.0:1	217
6.4 Conclusion	221
CHAPTER 7: General conclusion	226
APPENDIX A	231
APPENDIX B	243
APPENDIX C	248
APPENDIX D	267

LIST OF FIGURES

Figure 1.1. Schematic representation of <i>N</i> -linked and <i>O</i> -linked carbohydrates on glycoproteins. Man = mannose, Gal = galactose, NeuAc = N-acetylneuraminic acid. ^[30] [Reproduced with the permission of Ref. 30].	6
Figure 1.2. (a) Schematic representation of a Gram-positive (right) and a Gram-negative (left) bacteria. ^[46] (b) Representation of the Gram-positive (right) and Gram-negative (left) cell envelopes (WTA = wall teichoic acid, CAP = covalently attached protein, IMP = integral membrane protein, LP = lipoprotein, LPS = lipopolysaccharide, LTA = lipoteichoic acid, OMP = outer membrane protein. ^[45] [Reproduced with the permission of Ref. 45 and 46].	10
Figure 1.3. Molecular mechanisms of the generation of immune responses against protein-polysaccharide conjugate vaccines. ^[41] [Reproduced with the permission of Ref. 41].	13
Figure 1.4. Schematic representation of the mechanism of electrospray ionization. ^[83] [Reproduced with the permission of Ref. 83].	19
Figure 1.5. Schematic representation of the ionization of the matrix and the analyte in a MALDI source. ^[99] [Reproduced with the permission of Ref. 99].	21
Figure 1.6. Schematic representation of a quadrupole mass filter. ^[111] [Reproduced with the permission of Ref. 111].	24
Figure 1.7. Schematic representation of a linear MALDI-TOF-MS instrument used in the positive ion mode. ^[114] [Reproduced with the permission of Ref. 114].	26
Figure 1.8. Schematic representation of a QTOF. ^[118] [Reproduced with the permission of Ref. 118].	27

Figure 1.9. Schematic representation of the ion trap. ^[130] [Reproduced with the permission of Ref. 130].....	29
Figure 1.10. Schemaic representation of a cubic ICR cell. ^[133] [Reproduced with the permission of Ref. 133].....	30
Figure 1.11. Representation of the Thermo Fisher Orbitrap TM . ^[139] [Reproduced with the permission of Ref. 139].....	31
Figure 1.12. Tandem mass spectrometry principle. ^[149]	32
Figure 1.13. Oligonucleotides fragmentation nomenclature established by McLuckey and coworkers. ^[160]	35
Figure 1.14. Gas phase fragmentation nomenclature of a peptide. ^[170,171] [Reproduced with the permission of Ref. 170].....	37
Figure 1.15. Carbohydrate fragmentation nomenclature developed by Domon and Costello. ^[173] [Reproduced with the permission of Ref. 173].....	38
Figure 2.1. (a) Nomenclature for different oligonucleotide fragment types which includes the 3' or 5' terminus and (b) complementary fragment ions formation.(example: [a ₂ - B ₂] ^{k+} and w ₂ ^{x+}), example of internal fragment ion [pd(A)pf] ⁻ and [pd(A)] ⁻	65
Figure 2.2. Low-energy CID tandem mass spectra of the [M - 8H] ⁸⁻ anion at <i>m/z</i> 683.9881 of the oligomer <i>1</i>	74
Figure 2.3. Different specific product ions obtained during the CID-MS/MS analysis of the [M - 8H] ⁸⁻ ion at <i>m/z</i> 683.9881, for the following (a) oligomer <i>1</i> , (b) oligomer <i>2</i> and (c) oligomer <i>3</i>	76
Figure 2.4. Low-energy CID tandem mass spectra of the [M - 9H] ⁹⁻ anion at <i>m/z</i> 607.8775 of the oligomer <i>1</i>	78

Figure 2.5. Different specific product ions obtained during the CID-MS/MS analysis of the $[M - 9H]^{9-}$ ion at m/z 607.8775, for the following (a) oligomer 1, (b) oligomer 2 and (c) oligomer 3.....	79
Figure 2.6. Common product ions (isobarics and non isobarics) to at least two oligonucleotides obtained during the CID-MS/MS analysis of the $[M - 8H]^{8-}$ ion at m/z 683.9881, for the following (a) oligomer 1, (b) oligomer 2 and (c) oligomer 3.....	81
Figure 2.7. Common product ions (isobarics and non isobarics) to at least two oligonucleotides obtained during the CID-MS/MS analysis of the $[M - 9H]^{9-}$ ion at m/z 607.8775, for the following (a) oligomer 1, (b) oligomer 2 and (c) oligomer 3.....	82
Figure 3.1. (a) Schematic representation of the general structure of carbohydrate-spacer-squaric acid-protein constructs from oligosaccharide fragments of the <i>O</i> -polysaccharide of the <i>Vibrio cholerae</i> O1, serotype Ogawa and (b) MALDI-TOF/TOF-MS analysis of native BSA, and (c) hapten-BSA neoglycoconjugates with a hapten:protein ratio of 4.3.....	97
Figure 3.2. MALDI-TOF/TOF-MS spectra of the digested (a) BSA, and hapten-BSA conjugates with hapten:protein ratios of (b) 4.3, (c) 6.6 and (d) 13.2.....	104
Figure 3.3. MALDI-CID-TOF/TOF-MS/MS spectra of the glycosylated peptide (a) SLGK*VGTR at m/z 1330.72, (b) ALK*AWSVAR at m/z 1514.82 and (c) K*VPQVSTPTLVEVSR at m/z 2153.16.....	110
Figure 3.4. Hapten fragments observed during the the MALDI-TOF/TOF-MS/MS analysis of the SLGK*VGTR hapten-peptide conjugate at m/z 1330.72.....	111
Figure 3.5. MALDI-TOF/TOF-MS analysis of the neoglycoconjugates with hapten:BSA ratios 4:1 (a), 7:1 (b) and 13:1 (c), with the collision cell switched on.....	116

Figure 4.1. Schematic representation of the general structure of carbohydrate-protein constructs from oligosaccharide fragments of the antigenic tetrasaccharide side chain of the <i>B. anthracis</i> exosporium.....	128
Figure 4.2. General strategy applied for the mass spectrometry determination of the glycation sites on the hapten-BSA glycoconjugate.....	133
Figure 4.3. MALDI-MS analysis of the glycoconjugate trypsin digests (a) and GluC V8 endoproteinase digests (b).....	136
Figure 4.4. (a) MALDI-TOF/TOF-MS/MS spectra of the glycated peptide ALK*AWSVAR (Lys 235) at m/z 1951.0130. (b) Different product ions involving the fragmentation of the carbohydrate hapten observed during the MALDI-TOF/TOF-MS/MS analysis of the glycated peptide ALK*AWSVAR (Lys 235) at m/z 1951.0130.....	140
Figure 4.5. LC-QqTOF-MS/MS spectra of the tryptic glycated peptides (a) CCTK*PESER (Lys 463) at m/z 1058.4652 (+2), (b) CASIQK*FGER (Lys 228) at m/z 1073.0144 (+2) and GluC V8 digests (c) K*VTKCCTE (Lys 495) at m/z 987.9658 (+2) and (d) K*QEPERNE (Lys 117) at m/z 989.9697 (+2).....	147
Figure 4.6. (a) BSA sequence where the glycation sites are indicated by an asterisk (red = identified on tryptic digests, blue = identified on GluC V8 digests and red and underlined = identified on both tryptic and GluC V8 digests) and (b) 3D structure of the BSA. The glycated lysine residues are highlighted in red (Swiss-Pdb Viewer software).....	152
Figure 5.1. LC-ESI-QqTOF-MS/MS spectra of the tryptic glycopeptides (a) EK*VLTSAR (Lys 211) at m/z 752.4233 (+2), (b) GACLLPK*IETMR (Lys 204) at m/z	

951.5091 (+2), (c) VTK*CTESLVNR (Lys 498) at m/z 990.4813 (+2) and FK*DLGEEHFK (Lys 36) at m/z 881.9491 (+2).....	173
--	-----

Figure 5.2. Different product ions involving the fragmentation of the carbohydrate hapten observed during the LC-ESI-QqTOF-MS/MS analysis of the tryptic glycopeptide EK*VLTSSAR (Lys 211) at m/z 752.4233 (+2).....	174
---	-----

Figure 5.3. LC-ESI-QqTOF-MS/MS spectra of the glycopeptides obtained during the digestion of the neoglycoconjugates with the GluC V8 endoproteinase: (a) K*SHCIAE (Lys 309) at m/z 679.3320 (+2), (b) VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7646 (+3), (c) GPK*LVVSTQTALA (Lys 597) at m/z 899.5080 (+2) and VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 987.8313 (+3).....	180
--	-----

Figure 5.4. Neoglycoconjugate model sequences where the glycation sites are indicated by an asterisk (red = identified on tryptic digests, blue = identified on GluC V8 digests and red and underlined = identified on both tryptic and GluC V8 digests) for the neoglycoconjugates with a hapten-BSA ratio of (a) 4.3:1, (b) 6.6:1 and (c) 13.2:1.....	183
---	-----

Figure 5.5. 3-D structure of the neoglycoconjugate model vaccines. The glycated lysine residues are highlighted in red (Swiss-Pdb Viewer software) for the neoglycoconjugates with a hapten:BSA ratio of (a) 4.3:1, (b) 6.6:1 and (c) 13.2:1.....	184
--	-----

Figure 6.1. Schematic representation of the lactose-BSA glycoconjugate.....	201
--	-----

Figure 6.2. LC-QqTOF-MS/MS spectra of the tryptic glycated peptides (a) SLGK*VGTR (Lys 455) at m/z 697.3567 (+2), (b) LSQK*FPK (Lys 245) at m/z 712.3673 (+2), (c) EK*VLTSSAR (Lys 211) at m/z 783.8900 (+2) and (d) ALK*AWSVAR (Lys 235) at m/z 789.4125 (+2).....	208
---	-----

Figure 6.3. Product ions resulting from the tandem mass spectrometry analysis of the glyctated peptide SLGK*VGTR (Lys 455) at m/z 697.3567 (+2) extracted from the LC-ESI-Qq-TOF-MS/MS analysis of the tryptic digests of the hapten-BSA glycoconjugate with a lactose:BSA ratio of 5.1:1.....210

Figure 6.4. LC-QqTOF-MS/MS spectra of the GluC V8 digests glycated peptides (a) VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7418 (+3), (b) DKGACLLPK*IE (Lys 204) at m/z 910.4582 (+2), (c) VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463)) at m/z 1029.8189 (+3) and (d) DK*GACLLPK*IE (Lys 197, Lys 204) at m/z 1198.5604 (+2).....215

Figure 6.5. BSA sequence where the glycation sites are indicated by an asterisk (red = identified on tryptic digests, blue = identified on GluC V8 digests and red and underlined = identified on both tryptic and GluC V8 digests) for the neoglycoconjugates with a lactose:BSA ratio of (a) 5.1:1 and (b) 19.0:1.....216

Figure 6.6. 3-D structure of the lactose-BSA glycoconjugates with a lactose:BSA ratio of (a) 5.1:1 and (b) 19.0:1220

LIST OF TABLES

Table 2.1. Molecular and fragment ions obtained during the conventional ESI-QqTOF-MS analysis of the oligomers <i>I- 3</i>	71
Table 3.1. Molecular masses and calculated hapten:BSA ratios of BSA and hapten-BSA conjugates.....	102
Table 3.2. MALDI-TOF/TOF-MS peptide mapping of the digested BSA (after recalibration).....	106
Table 3.3. MALDI-TOF/TOF-MS peptide mapping of the digested neoglycoconjugate with a hapten:protein ratio of 4.3:1,after recalibration (glycated peptides bolted).....	107
Table 3.4. Molecular masses of BSA and hapten-BSA conjugates and their calculated hapten:BSA ratios, obtained during the MALDI-TOF-MS analysis of the glycoconjugate with the collision cell switched on.....	117
Table 4.1. Tryptic glycopeptides identified of the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the hapten:BSA glycoconjugate.....	146
Table 4.2. Glycopeptides identified in the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate digested with the endoproteinase GluC V8.....	150
Table 5.1. Tryptic glycopeptides identified during LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate with a hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1.....	170
Table 5.2. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide EK*VLTSSAR (Lys 211) at m/z 752.4233 (+2).....	175

Table 5.3. GluC V8 glycopeptide digests identified during LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate with a hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1	177
Table 5.4. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7646 (+3).....	181
Table 5.5. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 987.8313 (+3).....	190
Table 6.1. Molecular ions identified during the MALDI-TOF/TOF-MS analysis of the lactose-BSA glycoconjugates obtained with different reaction times.....	204
Table 6.2. Tryptic glycopeptides identified in the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the lactose-BSA glycoconjugates.....	207
Table 6.3. Glycopeptides identified in the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests of the lactose-BSA glycoconjugates.....	213

LIST OF ABBREVIATIONS

3-D: 3-Dimensional

α -CHCA: α -Cyano-4-hydroxycinnamic acid

A: Adenine

AA: Amino Acid

ACN: Acetonitrile

APPI: Atmospheric Pressure Photoionization

Asn: Asparagine

BSA: Bovine Serum Albumin

C: Cytosine

CAP: Covalently Attached Protein

CE: Capillary Electrophoresis

CI: Chemical Ionization

CID: Collision Induced Dissociation

CPS: Capsular Polysaccharide

Da: Dalton

ddNTP: Dideoxynucleotide triphosphate

DESI: Desorption Electrospray Ionization

DNA: Deoxyribonucleic Acid

DP: Declustering Potential

DTT: Dithiothreitol

ECD: Electron Capture Dissociation

EI: Electron Ionization

ETD: Electron Transfer Dissociation

ESI: Electrospray Ionization

DC: Direct Current

FA: Formic acid

FAB: Fast Atom bombardment

FP: Focusing Potential

FT-ICR: Fourier Transform Ion Cyclotron Resonance

FWHM: Full Width at Half Maximum

G: Guanine

Gal: Galactose

GalNAc: *N*-acetylgalactosamine

GC: Gas Chromatography

GlcNAc: *N*-acetylglucosamine

GS: Gas

HPLC: High performance liquid chromatography

ICR: Ion Cyclotron Resonance

IMP: Integral Membrane Protein

LC: Liquid Chromatography

LPS: Lipopolysaccharide

LTA: Lipoteichoic acid

Lys: Lysine

MALDI: Matrix Assisted Laser Desorption Ionization

Man: Mannose

MS: Mass Spectrometry

MS/MS: Tandem mass spectrometry

m/z : Mass-to-charge ratio

NCBI: National Center for Biotechnology Information

NeuAc: N-acetylneuraminic acid

NMR: Nuclear Magnetic Resonance

NTP: Nucleotide triphosphate

OMP: Outer Membrane Protein

Pa: Pascal

PCR: Polymerase chain reaction

PMF: Peptide mass fingerprint

ppm: part per million

Pro: Proline

Q: Quadrupole

QIT: Quadrupole ion trap

QTOF: Quadrupole Time Of Flight

RF: Radiofrequency

RNA: Ribonucleic acid

SELDI: Surface Enhanced Laser Desorption Ionization

Ser: Serine

T: Thymine

TFA: Trifluoroacetic acid

Thr: Threonine

TOF: Time of flight

U: Uracil

WTA: Wall Teichoic Acid

CHAPTER 1: Introduction

The scientific connotation of "biomolecule" denotes any molecule which is generated by a living organism. Biomolecules can be classified in two categories. The first is large polymeric biomolecules such as proteins, polysaccharides, lipids, and nucleic acids. The second is small molecules like the natural products, metabolites and secondary metabolites. The polymeric oligonucleotides, proteins, and glycoproteins are the main constituents of the large biomolecules group which are involved in the complex biochemical mechanisms of life. For example, attempting to understand the functions of biomolecules and their modification during the development of diseases is tedious, crucial and necessary for the understanding and design of new medical treatments. Therefore, researchers and medical practitioners are doing important work in developing new drugs and the necessary tools for the analysis and characterization of biomolecules, in a quality control manner. The chemical analysis of these new biomolecular drugs is necessary for following up the level and excretion of the drugs and finally for measuring the metabolites produced by the biological system. Accordingly, mass spectrometry (MS) has been proven to be a well suited technique for the analysis of complex molecules.

In this thesis, MS based methods have been applied for the analysis of oligonucleotides and carbohydrate-protein glycoconjugate vaccines.

Mass spectrometry has been extensively used for the understanding of Deoxyribonucleic acid (DNA) alterations which lead to several diseases by means of discovering genetic mutations as well as DNA adducts. It was found that the gas-phase fragmentation of oligonucleotides is challenging and its sequence determination is limited

by its size. This is why we proposed to study the gas phase fragmentation of isobaric oligonucleotides and to compare their fragmentation pathway.

The second category of biomolecules that we have analyzed corresponds to carbohydrate-protein vaccines. Several studies have been carried out to determine the carbohydrate-to-protein ratio of synthetic glycoconjugates. However, we are the first in the literature to have proposed a mass spectrometry-based method for the glycation sites determination in different neoglycoconjugates.

1.1. Biomolecules

1.1.1. Oligonucleotides

Nucleic acids are typically linear polymers of nucleotides. When composed of fifty or less nucleic acid bases, they are called oligonucleotides. The nucleic acids are constituents of DNA, ribonucleic acid (RNA) and the messenger ribonucleic acid (mRNA), which play essential roles in several fundamental biological processes. In 1953, Watson and Crick discovered the molecular structure of DNA, which permitted them to establish the "central dogma of molecular biology".^[1-3] The dogma states that the DNA is first transcribed to mRNA which is then translated to a sequence of amino acids, forming proteins. Moreover, the nucleotides contain the genetic information and are the first level of its transmission.^[3] They are considered as a key step for the organization and execution of protein synthesis, and also have a function of secondary messengers, metabolic regulators, and components of vitamins.^[3,4] Nucleotides are also involved in thermodynamically unfavourable enzyme-catalyzed reactions by working as high-energy intermediates. The nucleic acids are divided in two classes of heterocyclic nitrogenous

bases: the purine derivatives adenine (A) and guanine (G), and the pyrimidine derivatives uracil (U), thymine (T), and cytosine (C). A major challenge for the detection and analysis of nucleic acids in complex biological samples is their isolation, separation and purification.^[3] In addition, different studies also involve the purification and analysis of nucleotides, nucleosides and nucleic acids arising from hydrolyzed RNA and DNA.^[3]

Modified oligonucleotides

Since fifty years, it has been shown that the ability of many substances to bind to DNA is one of the causes of their mutagenic activity.^[3] Thus, carcinogenic substances have been extensively studied and the detection and measurement of carcinogen-DNA adducts has been found to be reliable for the determination of the human exposure to specific carcinogens and for evaluating their presence and incidence in the environment.^[3-8] The mass spectrometry (MS) analysis and fragmentation pathway determination for simple nucleosides and nucleotides is essential in order to understand the gas-phase fragmentation of more complex nucleic acids and oligonucleotides. McCloskey and coworkers developed extensively used MS techniques in nucleic acid research.^[3,4,9-11]

Synthetic oligonucleotides

The introduction of methods for the synthesis of oligonucleotides allowed the development of several applications, such as the design of DNA vaccines. Indeed, different DNA vaccines have been produced against infectious pathogens^[12] and cancer.^[13,14] Thus, the MS analysis of oligonucleotides became of primary importance since establishing the size, purity, and sequence of nucleic acids is a prerequisite to their use as molecular probes in biomedical science. Significant progress in the area of

accurate mass determination, sequencing, and study of noncovalent interactions has been made possible by the use of MS ionization techniques such as matrix-assisted-laser-desorption/ionization coupled to time-of-flight (MALDI-TOF) and electrospray ionization (ESI) coupled to conventional analyzers and tandem mass spectrometric (MS/MS) instruments.^[2]

1.1.2. Carbohydrates

It is well known that carbohydrates (sugars, saccharides, or glycans) are ever-present in the cells. Carbohydrates can be observed on all cell surfaces, in the nucleus and cytosol of eukaryotes, and within the extracellular matrix.^[15] The glycans are involved in all primordial biological processes such as cell synthesis, structure, and cell-cell interactions. They also play essential roles in disease.^[16] Thus, it is of outmost importance to understand the structure and function of all glycans within an organism.^[17] Thus, the study of the biological functions and structures of complex glycans expressed by cells, tissues, and organisms is known as "Glycomics". It has been predicted that the structural characterization of glycans will be the next step used to decode the molecular make-up of an organism.^[18]

The central dogma of biology is that all biological information flows from the DNA to the RNA and to the proteins. This dogma has allowed scientists to gain an understanding of biological organisms at a very fundamental level.^[19] When the mRNA is translated into proteins, a number of chemical modifications occur in the cells.^[20] The most important post-translational modification is the "glycosylation transfer reaction". This glycosylation reaction occurs when mono- or oligosaccharides are covalently attached through their anomeric C-1 position to other biomolecular aglycones such as the

proteins and/or the lipids forming new biomolecules such as the glycoproteins and the glycolipids.^[21] The glycosylation of proteins may alter the biological function of the protein and introduce a significant influence on all original biological function.^[22,23]

In the last century, glycoproteins have attracted a tremendous scientific interest due to their diverse structural and biological functions.^[22-25] This increased awareness was partly due to the reality that glycoproteins were discovered to be abundant in all living organisms. Indeed, glycoproteins are ubiquitous constituents of the cell wall; connective tissues such as collagen and are also found in gastrointestinal mucus secretions. Another facet indicating the biological importance of glycoproteins, is they are used as protective agents and lubricants and are also found abundantly in the blood plasma. The described diverse functions of glycoproteins are a straight result of their chemical and geometrical structures. Due to the tremendous biological importance of glycoproteins, it is important to reiterate, once more, that glycoproteins are formed by the glycosylation (covalent association) of one or more carbohydrate moieties to a peptide chain. This implies that glycoproteins consist of a polypeptide covalently bonded to a carbohydrate moiety also known as the "Glycan".

It is well known that glycans can exhibit enormous microheterogeneity and structural diversity. A glycosylated protein can exhibit several different spatial conformations,^[26] and could also present structural heterogeneity due to different "carbohydrate occupancy" also known as glycosylation sites. Thus, this constant challenge presented by the presence of the various factors of variations creates a very large structural diversity of glycan structures and make it a mammoth task to reveal the glycan structure and map their specific functions.

There are two major classes of glycoprotein structures which are known as the *N*-linked and the *O*-linked glycoproteins.^[27] *N*-linked carbohydrates are attached through *N*-acetylglucosamine (GlcNAc) and the amino acid asparagine (Asn). It was found that the *N*-linked amino acid consensus sequence is Asn-any amino acid (A.A.)-serine (Ser) or threonine (Thr). In addition, the middle amino acid can not be proline (Pro).^[27] *N*-linked oligosaccharides of glycoproteins are in most cases complex and branched. They are involved in the regulation of several cellular proteins, including enzymes and transcription factors.

The majority of *O*-linked carbohydrate covalent attachments to proteins engage a linkage between the monosaccharide *N*-acetylgalactosamine (GalNAc) and the amino acids Ser or Thr. At this time there is not an *O*-linked amino acid consensus sequence.^[28,29] In addition, they play a role in recognition, interaction, and enzyme regulation.

Figure 1.1 displays a schematic representation of *N*-linked and *O*-linked carbohydrates on glycoproteins.^[30]



Figure 1.1. Schematic representation of *N*-linked and *O*-linked carbohydrates on glycoproteins. Man = mannose, Gal = galactose, NeuAc = N-acetylneuraminic acid.^[30] [Reproduced with the permission of Ref. 30].

Carbohydrates' Role in Biological Function

It is well known and an obvious fact that the carbohydrate chains of glycoproteins play a role in the spatial conformation and structure of the polypeptide. Thus, in human immunoglobulins, the carbohydrate chain coils around one of the protein domains and prevents contact of this domain with the adjacent domain.^[31] The net result is that this adjacent domain could no longer perform its ordinary function. They are located throughout matrices and they operate as receptors on cell surfaces to other transported cells and proteins (collagen), together giving strength and support to a matrix.^[32] Glycoproteins can also cross link to other proteoglycan molecules and form complex ordered structures within cartilage tissue. In the nerve tissues, the abundant glycoproteins appear to be associated with synaptosomes, axons, and microsomes. In addition, it is well known that the glycoproteins thrombin, prothrombin, and fibrinogen are involved in the blood clotting mechanism.^[33]

Some bacteria possess glycoproteins (called s-layers) forming a membrane layer that envelop the outermost constituents of cell walls.^[33] Moreover, glycoproteins can act as bacterial flagella which are a construct of bundles of glycoproteins attached on the cell's surface.^[33] In the *Plantae* kingdom, the glycoproteins are involved in cell wall construction, tissue differentiation, embryogenesis, and sexual adhesion.^[33,34] Mucins are glycoproteins with a high molecular weight and are present on all the surfaces of the internal epithelial. They have the appearance of viscous gel and act as protecting agents to the epithelium from chemical, physical, and microbial damages.^[33,35] In addition to being involved in protection, mucins also have the capacity to function as lubricants.

In immunology, Kabat showed that many immunoglobulins were actually glycoproteins.^[36] He also showed that the interaction of blood group substances with antibodies was determined mainly by the presence of the glycoproteins on the erythrocytes. It was made known that adding or removing one monosaccharide from a blood group structure could dramatically alter the antigenicity and will alter the corresponding blood type. Thus, soluble immune mediators for instance helper, suppressor, and activator cells bind to glycoproteins located on the surface of their target cells. In effect, B and T cells contain surface glycoproteins that bind these circulating bacteria. Similarly, glycoproteins can direct phagocytosis.

1.1.3. Bacteria

It is well known that bacteria, viruses and other microbes are all surrounded by a capsular membrane formed usually of either complex polysaccharides or glycoproteins.^[37] These glycans are very important for promoting the host pathogen interactions. In the case of pathogenic Gram positive and/or negative bacteria, usually either the capsular polysaccharides (CPS) and/or the lipopolysaccharides (LPS) are implicated in all virulence factors which prevent phagocytosis and hence promote the bacterial colonization and persistence on mucosal surfaces.^[38,39] Therefore, it is well accepted that the foreign CPS and LPS are recognized by the different receptors of the innate immune system. This results in the production of cytokines, chemokines and cellular adhesion molecules.^[40] Accordingly, bacterial polysaccharides will induce an adaptive immune response. In the past century, bacterial glycan structures were employed successfully for the development of vaccines and immune modulators.^[41]

1.1.3.1. Gram negative bacteria

In 1884, Christian Gram invented a staining method (Gram Stain) which allows differentiating bacteria in two distinct groups. Gram-negative bacteria correspond to bacteria that do not retain crystal violet dye when tested with the staining method, contrary to the Gram-positive bacteria which are able to hold the stain.^[42] Moreover, Gram-negative and Gram-positive bacteria are also separated in different groups. As an example, the proteobacteria correspond to an important family of Gram-negative bacteria which incorporate *Escherichia coli* (*E. coli*), *Salmonella* and several other bacteria.^[43]

Structure of the Gram negative cell wall

In 1969, the development of the electron microscope allowed Glauert and Thornley to discover that the Gram-negative cell envelope is composed of three main layers: the peptidoglycan, the inner and the outer layers.^[44] Moreover, the outermost part of Gram-negative bacterial membrane is divided into inner and outer portions which contains a coat of LPS.^[45] Both inner and outer membranes are composed of phospholipids (mainly glycerol-phospholipids) and proteins. The peptidoglycan layer is located within the gelatinous material, called periplasm, which separates the two layers. The outer leaflet of the outer membrane is mainly composed of the amphiphilic LPS moieties that contain a lipid portion (lipid A) embedded into the membrane environment.

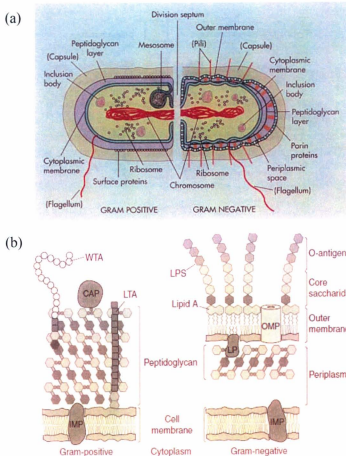


Figure 1.2. (a) Schematic representation of a Gram-positive (left) and a Gram-negative (right) bacteria.^[46] (b) Representation of the Gram-positive (left) and Gram-negative (right) cell envelopes (WTA = wall teichoic acid, CAP = covalently attached protein, IMP = integral membrane protein, LP = lipoprotein, LPS = lipopolysaccharide, LTA = lipoteichoic acid, OMP = outer membrane protein).^[45] [Reproduced with the permission of Ref. 45 and 46].

Figure 1.2 represents the configuration of Gram-positive and negative structures and also a representation of their cell envelope.^[46] The major difference between Gram-positive and Gram-negative bacteria is the absence of the outer membrane on Gram-positive bacteria. It has to be noted that the outer membrane plays a protective role for Gram-negative organisms and its presence allows the stabilization of the inner membrane.^[45]

Lipopolysaccharide (LPS)

Lipopolysaccharides (LPS) are present on the outer membranes in Gram-negative bacterial cells. Different methods have been developed for the isolation of LPS portions from bacteria, such as the aqueous phenol method in order to study their structure and composition.^[47] LPS correspond to amphiphilic macromolecules which are composed of an internal core oligosaccharide covalently linked to the lipid A, and also of an external core oligosaccharide (*O*-specific chain, or hydrophilic antigen). Lipopolysaccharides are characterized by endotoxic properties which are mainly due to the lipid A. Moreover, the *O*-specific chain consists of an arrangement of oligosaccharide units and its structure and composition are dependent on the bacterial serotype. Due to the fact that the polysaccharide moiety confers the immunological properties to the bacteria, several portions of LPS have been tested for therapeutic purposes, such as vaccines^[48] and anticancer drugs development.^[49]

LPS-derived vaccines or LPS-protein neoglycoconjugates

LPS derivatives have been mostly appreciated for their potential use as vaccine candidates. Bacterial LPS is the major contributor to the host immune reaction and hence, it has been successfully utilized for vaccination. The significance of LPS can be illustrated with the use of LPS antibodies which can protect against infection.^[50,51] LPS when injected alone is not immunogenic as it can escape recognition by the immune system due to its relatively small size.^[50] LPS conjugated to a protein carrier, however, can provoke the desired immunological reaction and can induce prolonged immunity. The first attempts to test protein conjugates as immunogens dates back to the late 1920s by Landsteiner's group.^[52,53] Shortly after, it was shown that bacterial carbohydrate-protein conjugates can induce strong antibody response.^[54] It is well known that the sugar portion of a carbohydrate-protein conjugate is always referred to as a hapten, which emerged from the work of Landsteiner's group.^[53,55] Figure 1.3 displays the immune response induced by protein-polysaccharide conjugate vaccines.^[41] First of all, the polysaccharide chain from the neoglycoconjugate vaccine cross links the B cell receptor, inducing the stimulation of these lymphocytes and the production of immunoglobulins. During the same time, the protein-polysaccharide conjugate vaccine is recognized by receptors located at the surface of polysaccharide specific B cells which process the protein carrier. Peptides specific to this protein are presented to carrier-peptide-specific T cells. As a consequence, T-cells are involved in the production of memory B cells and plasma B cells.

The covalent linkage between the LPS backbone and the protein carrier can be completed by either a direct linkage or with the aid of a suitable spacer. Many protein carriers have been assessed for the production of LPS-derived vaccines such as bacterial toxoids (e.g. tetanus toxoid),^[56] outer bacterial membrane proteins^[57,58] and bovine serum albumin (BSA).^[59] Regardless of the nature of the protein carrier, it is necessary to develop a neoglycoconjugate that is safe and capable of enhancing LPS immunogenicity. The first commercially available glycoconjugate vaccine was licensed in the late 1980s against *Haemophilus influenza* type b and this work was originally pioneered by Schneerson and co-workers.^[60] The infection has been virtually eliminated in countries where widespread vaccination among the general public has been executed.^[61]

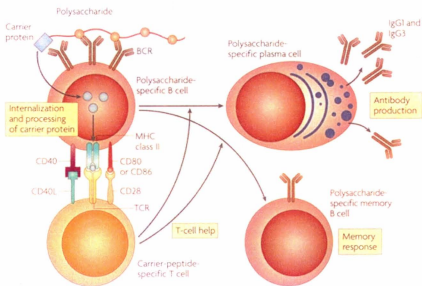


Figure 1.3. Molecular mechanisms of the generation of immune responses against protein-polysaccharide conjugate vaccines.^[41] [Reproduced with the permission of Ref. 41].

While many studies have focused on using natural LPS for the development of LPS vaccine conjugates, synthetic carbohydrate chains that mimic LPS epitopes (the part of an antigen recognized by the immune system) seem to be evolving as a promising approach for the production of LPS-protein conjugates. Pavol Kováč's group, for example, has been spearheading the production of hapten-BSA conjugates that bear synthetic mono-, tetra- and hexa-sugars which mimic the terminal part of the *O*-specific polysaccharide of the *Vibrio cholerae*, serotype Ogawa.^[62,63] Their novel conjugate (bearing the hexa-sugar) has shown protective immunity within mice.^[59] However, BSA-LPS derived from the *Vibrio cholerae*, serotype Inaba (despite structural similarities to Ogawa) failed to achieve this protective immunity.^[64] This has been attributed to the possibility that either the terminal sugars of Inaba are not protective epitopes or the delivery system was not optimized. The only difference between those two species lies within the terminal sugar which is characterized by a hydroxyl group at the 2 position in Inaba but a 2-*O*-methyl group in Ogawa. The investigators have successfully utilized surface-enhanced laser desorption/ionization time of flight (SELDI-TOF) MS to evaluate the degree of BSA-sugar conjugation and have presented very promising results with respect to controlling the reaction and the degree of substitution.^[62,65] These results, based on MS analysis, can have significant industrial impact during commercial production as it relates to quality control and quality assurance.

1.1.3.2. Gram positive bacteria

Gram-positive bacteria are characterized by their ability to be stained violet by Gram staining. They are also divided into numerous groups, including for example the cocci and the bacillus bacteria.^[66,67]

Structure of the Gram positive cell wall

Several differences can be observed on the cell envelope of Gram-positive and Gram-negative bacteria (Figure 1.2).^[46] First of all, in the Gram-positive bacteria, the outer membrane is not present. As a consequence, the layers of peptidoglycans that surround the bacteria is thicker than on Gram-negative bacteria.^[45] Moreover, the peptidoglycans in Gram-positive bacteria are linked to long anionic polymers called teichoic acids, which are composed of glycosyl phosphate, glycerol phosphate, or ribitol phosphate repeating units. Another major difference is the presence of covalently attached proteins at the surface of the Gram-positive bacteria.

Endospore

Some Gram-positive bacteria are able to develop an endospore in response to nutrient deprivation and to protect the bacteria from extreme stress (temperature, high UV irradiation, desiccation, chemical damage and enzymatic destruction).^[68] The endospore is composed of a core, a spore cell wall (or core wall), a cortex, a spore coat, and an exosporium. It has to be noted that the exosporium is a complex assembly of proteins, amino and neutral polysaccharides, lipids and ash.^[69] The exosporium is also the first part of the bacteria to be in contact with the host which confers to its constituents a major importance. It was also reported that spores of the following bacteria include specific glycoproteins at their surface layer: *Bacillus thuringiensis*,^[70] *Bacillus cereus*^[71] and *Bacillus anthracis*.^[72]

Carbohydrate-protein neoglycoconjugates vaccines

One of the most studied Gram-positive bacteria is the *Bacillus anthracis*.^[72] This bacteria causes anthrax disease (infection) mainly on herbivores and cattle. *Bacillus*

anthracis spores can also be used as biochemical warfare agents against humans and their inhalation are lethal within few days. Contrary to Gram-negative bacteria, Gram-positive bacteria do not possess LPSs.^[46] However, like most bacteria, *Bacillus anthracis* exhibits specific carbohydrates (on the surface of the spore) which can be used for the design of potential vaccines. Moreover, Sylvestre *et al.* found that the collagen-like spore surface glycoprotein of *Bacillus anthracis* (BclA) is extremely immunogenic and is part of the filaments of the hairy nap constituents.^[72] The identification of a tetrasaccharide antigen of the BclA allowed the synthesis of carbohydrate-protein vaccine models for the *Bacillus anthracis* bacterium.^[73]

1.1.3.3. Future perspectives in glycomics

It is obvious that scientists must decode the glycome. To decode the glycome, scientists must perform feats such as performing stereospecific carbohydrate synthesis, tedious structural characterization of the complex carbohydrates and associated glycoforms.

The current methods commonly employed to characterize the glycome include mainly mass spectrometry methods which are used to determine the structures of individual glycans located at the cell surface and also to identify sites of attachment of glycans on proteins.^[74] Nuclear magnetic resonance spectroscopy (NMR) is also used routinely to study the structure and function of glycan assemblies.^[75] Other biochemical methods such as lectin and antibody arrays and other higher throughput techniques can be used to identify and characterize the glycans used by the microbes which lead to infection.^[76] In addition, recently there have been reports discussing the development of

specific inhibitors for carbohydrate biosynthetic enzymes and carbohydrate binding proteins, to reveal their biological functions.

1.2. Mass spectrometry

Mass spectrometry is a technique that allows the detection of charged particles in the form of their mass-to-charge (m/z) ratio. This technique is used for the mass determination of analytes and also for their structure determination. Qualitative and quantitative information can be obtained with this technique. The sensibility, accuracy, low consumption and rapidity of analysis make MS a technique of choice for the biomolecules analysis. Mass spectrometry is employed in different fields such as proteomics,^[77] glycomics,^[78] metabolomics,^[79] lipidomics,^[80] and in oligonucleotides analysis.^[81]

A mass spectrometer is essentially composed of three main parts: an ionization source, a mass analyzer, and a detector.

1.2.1. Ion sources

The function of the ionization source is to generate charged particles. Different ionization sources have been developed, such as Fast Atom Bombardment (FAB), Electron Ionization (EI), Chemical Ionization (CI), ESI, Atmospheric Pressure Photoionization (APPI), MALDI, and more recently, Desorption Electrospray Ionization (DESI). A major consideration for the choice of an ionization source is the quantity of internal energy that can be transferred to the molecule during the ionization process and its physico-chemical properties.

Electrospray ionization and MALDI ionization techniques are qualified as “soft” because the analyte gains a minimum of internal energy during the ionization process, which lowers the fragmentation. The most commonly used soft ionization methods are Fast Atom Bombardment (FAB), Electrospray ionization (ESI) and Matrix Assisted Laser Desorption Ionization (MALDI).

1.2.1.1. Electrospray Ionization (ESI)

The principle of ESI is very simple: a liquid composed of the analyte and a polar solvent is passing through a capillary tube with a low flow rate ($1-10 \mu\text{l min}^{-1}$), and is submitted to a strong electric field in atmospheric pressure to create the electrospray.^[82] A potential difference of 2-5 kV between the capillary and the counter electrode is usually applied in order to create the electric field (Figure 1.4).^[83] On the end of the capillary tube, the liquid accumulates a charge at its surface, which breaks to produce charged droplets. Lord Rayleigh was the first to determine theoretically the maximum quantity of charge that a liquid droplet can carry (known as the Rayleigh limit).^[84] When the Coulomb forces reach the Rayleigh limit, the charged droplets break down to smaller fine charged droplets. Moreover, the dispersion of the spray is also obtained by an inert gas (usually nitrogen) which is injected coaxially to the capillary tube.

The formed droplets thus experience a curtain of heated inert gas in order to maximize the desolvation process. The charged analyte molecules are then accelerated toward the mass analyzer for detection.

During the electrospray experiment, the analyte molecules acquire a charge through proton transfer reactions between the electrospray solution and the analyte. However, electrochemical and photoelectrochemical reactions which take place at the electrode interfaces of the source have an influence on the ionization process.^[85]

The importance of ESI is due to its ability to form multiply charged ions. In the case of large biomolecules with a high molecular mass, such as proteins or oligonucleotides, the formation of multiply charged ions allows to obtain analyte molecules with a mass-to-charge ratio detectable with mass analyzers with a low nominal mass limit. In addition, typical mass spectra obtained by ESI-MS of a biomolecule correspond usually to a statistical distribution of multiply charged ions produced through protonation $[M + nH]^{n+}$, or deprotonation $[M - nH]^{n-}$.^[86] The connection of the ESI to a quadrupole mass analyzer allowed the development of the first mass spectrometer capable of analyzing proteins with a molecular mass up to 40 kDa.^[87,88] It has also to be noted that during the ESI, low or none fragment ions are observed.

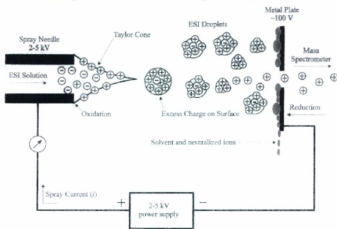


Figure 1.4. Schematic representation of the mechanism of electrospray ionization.^[83] [Reproduced with the permission of Ref. 83].

Another advantage of ESI is its ability to handle flow rates up to $1000\ \mu\text{L min}^{-1}$ which makes it connectable to high performance liquid chromatography (HPLC) systems.^[89]

1.2.1.2. Matrix Assisted Laser Desorption ionization (MALDI)

Matrix-assisted laser desorption/ionization is an extensively used ionization source coupled with a TOF mass analyzer for the molecular mass determination of biomolecules and proteins identification.^[90] For example, MALDI-TOF-MS was widely used for the analysis of oligonucleotides,^[91] oligosaccharides,^[92] proteins,^[90] glycoproteins^[92] and lipids.^[93]

Its high sensitivity, wide mass range, tolerance for salt and the fact that this method produces a low amount of fragments, allowed the use of MALDI for the direct analysis of biomolecules on tissue slices for MALDI imaging.^[94]

The principle of the MALDI experiment consists on the ionization of the analyte through a proton transfer reaction between a matrix and the analyte. First of all, the matrix solution, composed of a compound which is able to absorb the laser energy, is usually mixed with the analyte before spotting the mixture on a MALDI plate. However, different spotting methods have been experienced, such as quick and dirty, dried droplet, sandwich, seed-film layer and two-layer.^[95] The mixture is then left to dry at room temperature, resulting in a co-crystallization of the analyte with the matrix. Different matrices have been tested for the analysis of analytes of different nature: 2,5-dihydroxybenzoic acid for oligosaccharides, peptides and proteins,^[96] sinapic acid for proteins and α -cyano-4-hydroxycinnamic acid (α -CHCA) for a mixture of peptides.^[97,98] A laser beam (usually a nitrogen laser at 337 nm) is then applied on the spot of the

MALDI plate and causes the desorption and the ionization of the matrix and the analyte (Figure 1.5).^[99] It has to be noticed that the ionization mechanism is still unclear. However, most scientists believe that the ionization process results from three distinct processes: the excitation of the matrix molecules by the laser photons, the desorption of the analytes in the gas phase, and finally the ionization of the analyte in the gas phase.^[100,101] Moreover, the laser irradiation with a photon energy $h\nu$ causes the excitation of the matrix (MH), illustrated by the following equation: $MH + h\nu \rightarrow [MH]^*$. A portion of the energy carried by the excited matrix molecule is then transferred to the analyte. During the same time, the analyte and matrix molecules are desorbed to the gas phase and get ionized by a proton transfer reaction which occurs either during the collision of matrix molecules with the analyte molecule, or during an acid/base reaction.^[102]

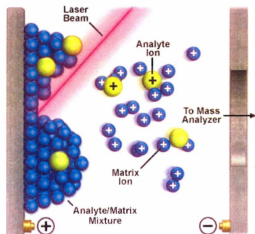


Figure 1.5. Schematic representation of the ionization of the matrix and the analyte in a MALDI source.^[99] [Reproduced with the permission of Ref. 99].

The advantage of the MALDI source is that it mainly provides low charged ions and thus facilitates the data interpretation. Recently, a major improvement of MALDI sources was the ability to operate under atmospheric pressure.^[103] Different atmospheric pressure sources, with the major advantage that they do not need a matrix, were then developed, including laser-induced acoustic desorption (LIAD) MS and desorption/ionization on silicon (AP-DIOS).^[104,105]

1.2.2. Mass analyzers

The major role of a mass analyzer is to separate different species (molecules, clusters or atoms) in their ionized form according to their m/z . Different mass analyzers were developed, but the widely used are the scanning analyzers where the separation of the ions is electronically-driven. This includes the quadrupole and the time-of-flight (TOF). Moreover, hybrid mass spectrometers were also developed which consists on the combination of different kind of mass analyzers. For example, the quadrupole-time-of-flight (QTOF) mass spectrometer is widely used in proteomics. Another type of mass analyzer was also developed: the trapping instruments. In these mass analyzers, the ions are trapped and an ion with a specific m/z can be selected (the other ions are expelled from the mass analyzer) in order to carry out different experiments, such as collision induced dissociation (CID) by introducing a neutral gas in the mass analyzer. The mainly used ion trap mass analyzers are the quadrupole ion trap, the orbitrap and the Fourier transform ion cyclotron resonance (FTICR).

The resolution, sensitivity, mass range and mass accuracy, which are the main characteristics to consider during the analysis, differ from one analyzer to another.

1.2.2.1. Quadrupole mass filter

The quadrupole mass filter is an extensively used mass analyzer and it has also been widely used for 30 years.^[106] This analyzer is frequently associated with an on-line separation technique, such as liquid chromatography (LC),^[107] gas chromatography (GC)^[108] and capillary electrophoresis (CE).^[109] However, its low resolution do not make the quadrupole a mass analyzer of choice for molecules characterization.^[110]

The principle of operation of this mass analyzer is based on the trajectory of the ions inside the quadrupole chamber. Basically, the quadrupole is composed of four parallel metal rods electronically connected (Figure 1.6).^[106,111] A direct current (DC) and a radiofrequency (RF) potential are simultaneously applied between one pair of rods and the other, which cause the oscillation of the ions between the rods. By changing the values of the DC and the RF, only one ion of a certain m/z value will have a stable trajectory in the quadrupole and reach the detector (these are resonant ions), while the ions with a different m/z value will collide with the cylinders and be lost (these are non resonant ions). The electric field applied on the rods is described by the following formula: $\Phi_0 = \pm (U \pm V \cos \omega t)$, where Φ_0 represents the voltage applied to the rods, U is the DC voltage, V is the RF voltage amplitude and ω is the frequency. Additionally, inside the quadrupole mass analyzer, the trajectory of the ions will follow the equation of Mathieu and will be separated according to their m/z . The quadrupole mass analyzer presents several advantages: excellent transmission efficiency, suitability for GC, LC and CE, low cost, robustness, easy to use. However, the mass range of the quadrupole (less

than 4000 Da), its low resolution and the need to couple more than two quadrupoles to achieve tandem mass spectrometry analysis are the major limitations of this analyzer.

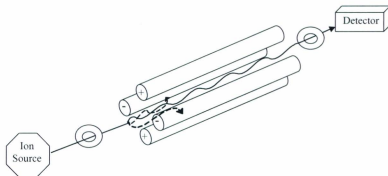


Figure 1.6. Schematic representation of a quadrupole mass filter.^[111] [Reproduced with the permission of Ref. 111].

1.2.2.2. Time of flight (TOF) mass analyzer

In 1946, the TOF mass analyzer was introduced for the first time by Stephens^[112] and the first linear TOF commercial instrument was developed by Wiley and McLaren in 1955.^[113] The TOF analyzer is well suited for pulsed ionization sources, such as the laser desorption ionization. Figure 1.7 shows the scheme of a linear MALDI-TOF mass spectrometer.^[114] Its principle is very simple: it consists in measuring the time it takes for an ion to traverse a field-free-time-of-flight tube. First of all, after the ionization process, the ion leaves the ionization source with a charge $q = Ze$ (Z = charge number, e = elementary charge) and a mass m . The ion has a kinetic energy $E_c = 1/2mv^2 = qVs$ (v = velocity of the ion). Thus, the time it takes for the ion to traverse the field-free region with a length L is measured and directly related to the m/z value of the ion through the kinetic energy equation previously mentioned.

An important parameter to be considered for a mass analyzer is its resolution. The TOF analyzer has a resolution which is dependent on the mass of the analyte. It is well known that the ions of the same m/z value have slightly different kinetic energies when they enter the TOF mass analyzer.^[115] This is the cause of a limitation of the resolution for the TOF analyzers. At higher masses, the resolution of the TOF is limited. Moreover, a reflectron was introduced in the TOF mass analyzer in order to improve the resolution. The reflectron is composed of a series of charged ring electrodes, in which the first ring has the lowest potential and the last ring the highest. Thus, the electrostatic field increases from the first ring to the last one.^[116] The reflectron is disposed at the end of the TOF analyzer. After traversing the TOF, the ions experience the increasing electrostatic field in the reflectron. They are then slowed down and stopped before being reflected and accelerated in the opposite direction. Moreover, the faster ions of a given m/z value have a higher kinetic energy and thus travel further into the reflectron before being reflected than the slower ions. Then the path followed by the faster ions is longer than the one followed by the slower ones. This phenomenon allows the slower ions to catch up with the faster ones and these ions arrive to the detector at the same time.

Different applications of the MALDI-TOF mass spectrometer were reported, such as for biomolecules and also for synthetic polymers and polymer/biomolecule conjugates.^[117]

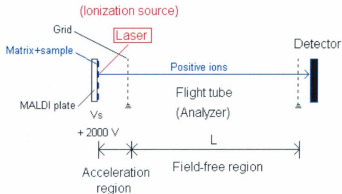


Figure 1.7. Schematic representation of a linear MALDI-TOF-MS instrument used in the positive ion mode.^[114] [Reproduced with the permission of Ref. 114].

1.2.2.3. Hybrid quadrupole/time-of-flight (QTOF) mass spectrometer

The hybrid mass spectrometers are composed of two or more mass analyzers of different types. They are usually designed to perform tandem mass spectrometry experiments. Figure 1.8 displays a schematic representation of a QTOF mass spectrometer.^[118]

The QTOF mass spectrometer consists on a quadrupole mass analyzer (Q) used in conjunction with an orthogonal TOF. In the QTOF mass spectrometer, the quadrupole is usually used to select the ion of interest while the TOF analyzer acquires the mass spectrum. However, when the quadrupole is operated with only the RF potentials applied to the rods, all the ions emerge from the quadrupole and are separated in the TOF analyzer.^[118,119]

Different ionization sources can be used in conjunction with the hybrid QTOF instruments, such as ESI,^[120] APPI^[121] and MALDI.^[122]

The QTOF mass spectrometer has different advantages. The ability to use an electrospray ionization source with the mass spectrometer permits to lower the internal energy transferred to the ion and thus minimizes the fragmentation of the molecular ions. Consequently, the analysis of large biomolecules and their mass determination has been achieved using ESI-QTOF MS.^[118-120,123] The second main advantage of this hybrid instrument is its high sensitivity and also its linear mass scale to 10,000.

The QTOF mass spectrometer is generally used with an electrospray inlet which allows the coupling with a LC. One major application of LC-ESI-QTOF MS is the proteins identification and sequencing.^[124-126]

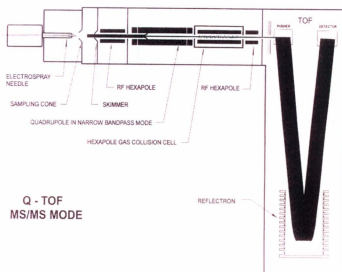


Figure 1.8. Schematic representation of a QTOF.^[118] [Reproduced with the permission of Ref. 118].

1.2.2.4. Trapping mass analyzers

The trapping mass analyzers correspond to analyzers where the ions can be retained (or trapped) for a certain period of time. Thus, ion/molecule reactions can be studied with ion trap analyzers. The main advantage of these mass analyzers is also their ability to perform consecutive tandem MS experiments.

1.2.2.4.1. Quadrupole Ion trap

The quadrupole ion trap (QIT) exists in two different forms: a 3D configuration, also known as a Paul trap, and a linear shape (linear QIT).

The Paul trap mass analyzer is composed of three hyperbolic electrodes, a central electrode and also two endcap electrodes (Figure 1.9).^[127-130] A DC current and a RF is applied between the electrodes and the resulting forces allow to trap the ions between the three hyperbolic electrodes. The ions are excited by the oscillation of the RF between the two endcap electrodes, and are directed inside the ion trap by applying a driving voltage to the ring electrode. Thus, by applying different voltages, the ions can be either trapped or expelled from the ion trap.

The linear quadrupole ion trap (linear QIT) mass analyzer is composed of a set of parallel rods built up in a square configuration. As for the quadrupole mass analyzer, a DC current and a RF is applied between two pairs of opposite rods and the ions are then trapped inside the linear QIT. In addition, an electric potential on the end of the electrodes allows the ions to be trapped axially in the ion trap.^[131] The motion of the ions between the parallel rods is described by the Mathieu Equations.

In general, the quadrupole ion trap presents the advantages to have a high sensitivity and a good resolution. However, its inconvenience resides in its low mass accuracy.^[130]

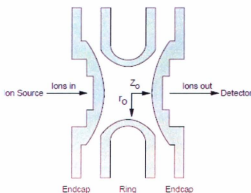


Figure 1.9. Schematic representation of the ion trap.^[130] [Reproduced with the permission of Ref. 130].

1.2.2.4.2. Fourier Transform Ion Cyclotron Resonance (FTICR)

In the FTICR MS, the ions experience a magnetic field and their mass m/z is determined based on their cyclotron frequency.^[132] The ion cyclotron resonance (ICR) cell is generally composed of a pair of trapping plates, a pair of excitation plates and a pair of detection plates (Figure 1.10).^[133] The ions can be trapped and stored under a magnetic field with electric trapping plates, also called a Penning trap. The application of an oscillating electric field perpendicular to the magnetic field effects excitation of the ions to a higher cyclotron radius, and concentrates them in packets and making them move in phase. When the ions pass close to the detection plates, they are detected as an image current and the signal is named a free induction decay which is composed of a

superposition of sine waves. The signal is thus recorded in a time domain and mass spectrum is the result of the Fourier transform of this signal.

The main advantage of the FTICR is its high resolution and one application of this technique is the molecule composition determination based on their accurate mass.^[134] However, FTICR-MS has also been applied for the analysis of proteins, glycoproteins and oligonucleotides.^[135-137]

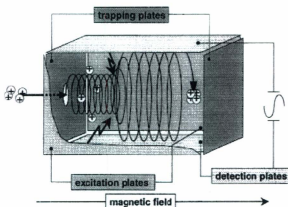


Figure 1.10. Schematic representation of a cubic ICR cell.^[133] [Reproduced with the permission of Ref. 133].

1.2.2.4.3. Orbitrap

In 2000, Alexander Makarov introduced the Orbitrap.^[138] This mass analyzer (Figure 1.11) is composed of a spindle-like-shaped central electrode (a), an outer electrode (b) and an insulating ceramic ring (c).^[139] After ionization, the ions are injected between the central and the outer electrodes and experience an electric field applied tangentially to the electrodes. The ions undergo an electrostatic attraction to the central electrode which is balanced by centrifugal forces, causing the ions to be trapped between

the two electrodes. Thus, the ions have a cyclic motion around the central electrode and also move on the axis of this electrode. The oscillations of an ion of a specific m/z along the central spindle create a signal in a similar way to the FTICR-MS, which is then converted through the Fourier transformation to a m/z value.^[139,140]

The resolving power (maximum of 200,000) and the mass accuracy (1-2 ppm) confer to this trapping instrument a real advantage and several applications have been reported such as the analysis of proteins,^[141] glycoproteins,^[142] metabolites,^[143] and oligonucleotides.^[144] In addition, a hybrid instrument has been developed which consists on the association of a linear quadrupole ion trap to the Orbitrap (LTQ-Orbitrap).^[145]

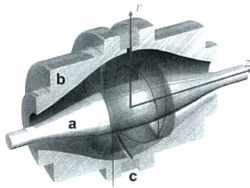


Figure 1.11. Representation of the Thermo Fisher OrbitrapTM.^[139] [Reproduced with the permission of Ref. 133].

1.2.3. Ions in the gas phase

1.2.3.1. Gas phase ions formation and their fragmentation

The formation of gas phase ions is occurring in the ionization source. It was observed that during the ionization process, the energy gained by the ionized molecule plays a role in the formation of in-source fragment ions.

1.2.3.2. Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) consists on multistage analyses.^[146] The principle is based on the isolation of a precursor ion which is induced to fragment. The resulting product ions are then detected.^[147] Tandem mass spectrometry has been extensively used over the past 10 years for the structure characterization of unknowns.^[148] There are two different MS/MS: tandem in space and tandem in time instruments.

In tandem in space, the mass spectrometer is composed of at least two analyzers and a collision cell. After the ionization process, a precursor ion is selected in the first analyzer and is introduced in the collision cell (Figure 1.12). The collisions of the precursor ion with the neutral gas inside the collision cell allowed the formation of product ions which are then separated in the second mass analyzer and detected.^[146,149,150] This procedure is also known as CID. Tandem in space MS is commonly used with the triple quadrupole mass spectrometer and also with hybrid instruments, such as the QTOF.

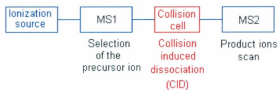


Figure 1.12. Tandem mass spectrometry principle.^[149]

For the tandem MS in time, the isolation of the precursor ion, its fragmentation and the detection of the product ions take place in an ion trap mass analyzer where the ions can be held (or trapped) for a specific time.^[150] The QIT, the FTICR and the Orbitrap can be used for this experiment.

In addition, different activation methods which are not available for instruments which perform tandem in space have been developed for the trapping instruments. For example, the electron capture dissociation (ECD) has been introduced for FTICR-MS.^[151] It consists on introduction of low energy electrons in the ICR cell containing multiply charged ions in the gas phase. The interaction between the charged ions and the electrons produces odd-electron ions which undergo dissociation to form different type of fragments. In the case of proteins and peptides, the ECD produces mainly c- and z-ions.^[151] Electron capture dissociation has been proven to be well suited for the analysis of glycopeptides because of its ability to form exclusively c- and z-ions, while the carbohydrate portion of the glycopeptide does not experience fragmentations.^[152] In the Orbitrap instrument, radical anions (anthracene or azobenzene) are used to transfer electrons to the charged molecules, and the following fragmentation process is similar to the one observed when used ECD in a FTICR-MS.^[153]

1.3. Mass spectrometry analysis of biomolecules

1.3.1. Mass spectrometry analysis of oligonucleotides

Deoxyribonucleic acid sequencing allows getting the sequence of nucleotides in a particular gene. Different strategies have been developed for the MS analysis and sequencing of oligonucleotides.

1.3.1.1. Sanger sequencing

In 1975, the group of Sanger introduced a new DNA sequencing method which is also known as the dideoxy sequencing method.^[154] Basically, a DNA portion is amplified using the polymerase chain reaction (PCR). However, in addition to nucleotides triphosphates (NTP), dideoxynucleotide triphosphates (ddNTP) are also integrated in the

DNA sequence which stops further additions of nucleotides. This will result in the production of DNA sequences of different lengths. The MALDI-TOF-MS analysis of the mixture of DNA strands of different size allows the sequencing of the DNA target through the comparison of the molecular ion masses. Indeed, by calculating the mass difference of two closest molecular ions, we can identify the nucleotide which correspond to this difference.^[155]

1.3.1.2. Ladder sequencing

In the Ladder sequencing method, the target DNA sequence is first amplified using PCR, and the resulting oligonucleotides are digested by means of an exonuclease enzyme (snake venom phosphodiesterase for example).^[156] After digestion, we generally obtain oligonucleotides of different size. The MALDI-TOF-MS analysis of the mixture allows determining the sequence of the oligonucleotide by identifying the nucleotides difference between two closest molecular ions.

1.3.1.3. Gas phase fragmentation of oligonucleotides

Several studies have been carried out on the gas phase fragmentation of oligonucleotides.^[157-159] McLuckey and coworkers proposed a nomenclature for the gas-phase fragmentation of oligonucleotides (Figure 1.13), which is based on a previously established nomenclature for the fragmentation of peptides.^[157] The fragment ions which include the 5' termini of the oligonucleotide are designed as a, b, c and d, while those including the 3' termini are named w, x, y and z. The same group also observed that during the CID-MS/MS analysis of oligonucleotides in the negative ion mode, the following product ions are formed: (a - B), w, y and (d - H₂O).

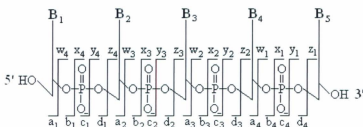


Figure 1.13. Oligonucleotides fragmentation nomenclature established by McLuckey and coworkers (B_1 , B_2 , B_3 and B_4 correspond to nucleic bases).^[157]

In 1999, Rozenski introduced an online software called Mongo Oligo Mass Calculator v2.06, which can predict the different possible product ions (($a-B$), w , y and ($d-H_2O$)) of an oligonucleotide based on its sequence.^[160] The software is also able to predict the series of multi-charged molecular ions that we may observe for the ESI-MS analysis of an oligonucleotide of known sequence.

Moreover, in an attempt to facilitate the oligonucleotide sequencing, Rozenski developed an interactive program called Simple Oligonucleotide Sequencer (SOS), for *ab initio* oligonucleotide sequencing by MS.^[161] The software uses the mass spectra and the sequencing of the oligonucleotide is carried out residue by residue. The following product ions can be identified in the negative ion mode: ($a-B$), w , y and ($d-H_2O$) which allows the construction of oligonucleotide ladders in both the $3' \rightarrow 5'$ directions and the $5' \rightarrow 3'$ directions. The comparison of the ladders can then be carried out for homology and overlap. The advantage of the software is the possibility to include modifications on bases, sugars or backbones. However, the disadvantage of this program is that it cannot be extend to oligonucleotides longer than *mer*-20 level.

1.3.2. Mass spectrometry analysis of proteins and peptides

Different strategies have been developed for the MS analysis, identification and sequencing of proteins.

“Shotgun proteomics” consists on the use of the “bottom-up approach” for the identification of proteins.^[162-164] First, the proteins are digested using a protease, such as trypsin. The resulting complex mixture of peptides is then separated using LC-MS/MS. The proteins are then identified by comparison of the MS data to theoretical peptide MS data calculated from a genomic or proteomic database (example: National Center for Biotechnology Information, NCBI) using search engines such as Mascot and Sequest.^[165-167] However, when the mixture of peptides is analyzed by MALDI-TOF-MS, the identification of the protein is carried out using the same search engines and is called peptide mass fingerprinting.^[162]

Another strategy is successfully applied which is the “top-down approach”. In this technique, the entire protein is analyzed.^[168] Molecular ions of the proteins are produced by electrospray and one of them is selected and subjected to gas phase fragmentation. In top-down proteomics it is necessary to work with high resolution mass analyzers. Indeed, one challenge with this method is the generation of product ion masses from multiply charged protein precursor ions. The interpretation of the tandem mass spectra allows the sequencing of the protein. The data can also be uploaded in search engines such as Mascot for the identification of the protein.

However, some proteins with unknown sequence or with post-translational modifications are difficult to identify and to sequence. The protein sequence is then determined with minimal support from genomic data, by *de novo* protein sequencing.

Both of the previously mentioned methods were applied for the *de novo* sequencing of proteins.^[169] Basically, the MS/MS spectra obtained by the “bottom-up” or “top-down” approaches are used for the determination of the protein sequence.

A gas-phase peptide fragmentation nomenclature was introduced by Roepstorff and Fohlman and lately modified by Johnson and coworkers (Figure 1.14).^[170,171] The fragment ions including the N-terminus of the peptide are called a, b and c, while the x, y and z fragment ions include the C-terminus.

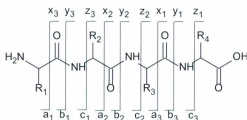


Figure 1.14. Gas phase fragmentation nomenclature of a peptide.^[170,171] [Reproduced with the permission of Ref. 170].

1.3.3. Mass spectrometry analysis of glycoconjugates

Glycoconjugates correspond to biomolecules such as proteins, peptides or lipid aglycones which are covalently linked to glycosyl chains of oligosaccharides or polysaccharides. Several MS methods have been applied for the structure determination of carbohydrates and glycoconjugates.^[172-174]

Domon and Costello were the first to report the CID-MS/MS analysis of protonated complex glycoconjugate molecules.^[172,174] They also introduced a universal nomenclature for the fragmentation of glycoconjugates.^[172] Figure 1.15 illustrates the possible product ions which can be obtained during the CID-MS/MS analysis of a

ganglioside, according to the nomenclature established by Domon and Costello. The A, B and C product ions correspond to ions that include the non-reducing end of the oligosaccharide, while the X, Y and Z product ions incorporate the aglycon (reducing end of the oligosaccharide). It has to be noted that X and A product ions are produced by the fragmentation of the glycosidic ring, and the ring bond which is broken to produce the ions is noted on A and X symbols.

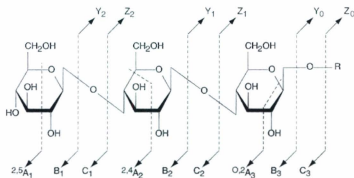


Figure 1.15. Carbohydrate fragmentation nomenclature developed by Domon and Costello.^[172] [Reproduced with the permission of Ref. 172].

1.4. Co-authorship statement

All experimental work presented in Chapters 2, 3, 4, 5 and 6 was performed by the principal author (Farid Jahouh), except the MALDI-TOF-MS experiments in Chapter 3 which were carried out by Donatella Aiello and Anna Napoli (group of Professor G. Sindona, Department of Chemistry, University of Calabria, Arcavacata di Rende (CS), Italy).

The carbohydrate hapten-BSA glycoconjugates (Chapters 3-6) were synthesized by the group of Professor Dr. P. Kováč (NIDDK, LBC, National Institutes of Health, Bethesda, MD 20892-0815, USA).

The group of Dr. P. Martin (Unité de Catalyse et de Chimie du Solide, site de l'Artois - UMR CNRS 8181, I.U.T. de Béthune, Département Chimie, 1230 rue de l'Université, BP819, 62408 Béthune cedex, France) and Dr. M. Nashed (Department of Chemistry, Faculty of Science, Alexandria University Alexandria, Egypt) contributed for the interpretation of the data in chapter 2.

The interpretation and discussion of the data were carried out by the principal author for chapters 3-6. The preparation of the draft manuscripts (based on chapters 2-6) was also achieved by the principal author, corrected by the supervisor, after extensive discussion with the co-authors.

References

- [1] J. D. Watson, F. H. C. Crick. Molecular structure of nucleic acids. A structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737.
- [2] F. H. C. Crick. Central dogma of molecular biology. *Nature* **1970**, *227*, 561.
- [3] J. H. Banoub, J. Miller-Banoub, F. Jahouh, N. Joly, P. Martin. Overview of recent developments in the mass spectrometry of nucleic acid and constituents. In *Mass Spectrometry of Nucleosides and Nucleic Acids*. J. H. Banoub, P. A. Limbach (Eds) CRC Press. **2010**, p. 3.
- [4] J. H. Banoub, S. Combden, J. Miller-Banoub, G. Sheppard, H. Hodder. Structural Characterization of Intact Covalently Linked DNA Adducts by Electrospray Mass Spectrometry. *Nucleosides and Nucleotides*. **1999**, *18*, 2751-2768.
- [5] K. Randerath, E. Randerath, H. P. Agrawal, R. C. Gupta, M. E. Schurdak, M. V. Reddy. Postlabeling methods for carcinogen-DNA adduct analysis. *Environ. Health Perspect.* **1985**, *62*, 57.
- [6] M. V. Reddy, R. C. Gupta, E. Randerath, K. Randerath. 32P-postlabeling test for covalent DNA binding of chemicals in vivo: application to a variety of aromatic carcinogens and methylating agents. *Carcinogenesis*. **1984**, *5*, 231.
- [7] K. Gorlewska-Roberts, B. Green, M. Fares, C. B. Ambrosone, F. F. Kadlubar. Carcinogen-DNA adducts in human breast epithelial cells. *Environ. Mol. Mutagen.* **2002**, *39*, 184.
- [8] P. B. Farmer, G. M. A. Sweetman. Mass spectrometric detection of carcinogen adducts. *J. Mass Spectrom.* **1995**, *30*, 1369.

- [9] J. A. McCloskey. Techniques for the structure elucidation of complex nucleosides by mass spectrometry, in *Proc. Th. Int. Roundtable on nucleosides, nucleotides and their biological application*. F. C. Alderweindelt, E. L. Esmans (Eds.) University of Antwerp, **1982**, p.47.
- [10] J. A. McCloskey. Mass spectrometry of nucleic acid constituents and related compounds, in *Mass Spectrometry in the Health and Life Sciences*, A. L. Burlingame, N. Castagnoli, (Eds.) Elsevier, Amsterdam, **1985**, p. 521.
- [11] J. A. McCloskey. Experimental approaches to the characterization of nucleic acid constituents by mass spectrometry, in *Mass Spectrometry in Biomedical Research*. S. J. Gaskell (Ed.) Wiley, New York, **1986**, p. 75.
- [12] R. E. Tascon, M. J. Colston, S. Ragno, E. Stavropoulos, D. Gregory, D. B. Lowrie. Vaccination against tuberculosis by DNA injection. *Nat. Med.* **1996**, *2*, 888.
- [13] F. K. Stevenson, C. J. Jr. Link, A. Traynor, H. Yu, M. Corr. DNA vaccination against multiple myeloma. *Semin. Hematol.* **1999**, *36*, 38.
- [14] A. Waisman, P. J. Ruiz, D. L. Hirschberg, A. Gelman, J. R. Oksenberg, S. Brocke, F. Mor, I. R. Cohen, L. Steinman. Suppressive vaccination with DNA encoding a variable region gene of the T-cell receptor prevents autoimmune encephalomyelitis and activates Th2 immunity. *Nat. Med.* **1996**, *2*, 899.
- [15] B. K. Brandley, R. L. Schnaar. Cell-surface carbohydrates in cell recognition and response. *J. Leukoc. Biol.* **1986**, *40*, 97.
- [16] T. Willer, M. C. Valero, W. Tanner, J. Cruces, S. Strahl. O-mannosyl glycans: from yeast to novel associations with human disease. *Curr. Opin. Struct. Biol.* **2003**, *13*, 621.

- [17] A. Varki, J. B. Lowe. Biological roles of glycans. In *Essentials of Glycobiology*. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler (Eds). Cold Spring Harbor Laboratory Press (NY), 2nd edition, **2009**, p. 75.
- [18] R. Raman, S. Raguram, R. Sasisekharan. Informatics concepts to decode structure-function relationships of glycosaminoglycans, in *Bioinformatics for Glycobiology and glycomics: an introduction*. C. W. Von der Lieth, T. Lütke, M. Frank (Eds). John Wiley & Sons, Ltd, Chichester, UK. **2009**, p. 3 .
- [19] F. Crick. Central dogma of molecular biology. *Nature* **1970**, 227, 561.
- [20] N. Malys, J. E. G. McCarthy. Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* **2010**, 68, 991.
- [21] R. G. Krishna, F. Wold. Post-translational modification of proteins, in *Advances in Enzymology and Related Areas of Molecular Biology*. A. Meister (Ed.), John Wiley & Sons, Inc., Hoboken, NJ, USA. Vol. 67, **2006**.
- [22] S. A. Brooks. Protein glycosylation in diverse cell systems: implications for modification and analysis of recombinant proteins. *Expert Rev. Proteomics* **2006**, 3, 345.
- [23] B. Dutta, R. Ain, P. B. Seshagiri, A. A. Karande. Differential influence of recombinant non-glycosylated and glycosylated glycodeclin on human sperm function: comparative studies with hamster spermatozoa. *Reprod. Fertil. Dev.* **2001**, 13, 111.
- [24] T. Muramatsu. Cell surface glycoproteins: biochemical, immunological and molecular biological studies. *Nagoya J. Med. Sci.* **1994**, 57, 95.
- [25] K. Olden, B. A. Bernard, M. J. Humphries, T. K. Yeo, K. T. Yeo, S. L. White, S. A. Newton, H. C. Bauer, J. B. Parent. Function of glycoprotein glycans. *Trends Biochem. Sci.* **1985**, 10, 78.

- [26] J. G. Beeley. Peptide chain conformation and the glycosylation of glycoproteins. *Biochem. Biophys. Res. Commun.* **1977**, 76, 1051.
- [27] N. L. Wilson, B. L. Schulz, N. G. Karlsson, N. H. Packer. Sequential analysis of N- and O-linked glycosylation of 2D-PAGE separated glycoproteins. *J. Proteome Res.* **2002**, 1, 521.
- [28] C. K. Mathews, K. E. Van Holde. *Biochemistry*. The Benjamin/Cummings Publishing Company, Inc.: Redwood City, **1990**, p. 291.
- [29] G. E. Schulz, R. H. Schirmer. *Principles of Protein Structure*. Springer-Verlag: New York, **1979**, p. 228.
- [30] T. Schwientek, 2010: <http://www.uni-koeln.de/med-fak/biochemie/staff/schwientek/index.e.shtml>
- [31] C. Chothia, E. Y. Jones. The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* **1997**, 66, 823.
- [32] R. Kornfeld, S. Kornfeld. Assembly of Asparagine-linked oligosaccharide. *Annual Rev. Biochem.* **1985**, 54, 631.
- [33] (a) A. Gottschalk. Glycoproteins. Their composition, structure, and function. Elsevier Publishing Company: New York, **1972**, p. 20. (b) J. Montreuil, J. F. G. Vliegenthart, H. Schachter. Glycoproteins. Elsevier: Amsterdam, 1995. Qiu Zhiyong, and Frank Tufaro and Shirley Gillam. "Brefeldin A and Monensin Arrest Cell Surface Expression of Membrane Glycoproteins and Release of Rubella Virus". *J. Gen. Virol.* 76 (1995): 885-863.

- [34] Q. Zhiyong, F. Tufaro, S. Gillam. Brefeldin A and monensin arrest cell surface expression of membrane glycoproteins and release of Rubella virus. *J. Gen. Virol.* **1995**, 76, 855.
- [35] A. Allen, G. Flemstrom. Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin. *Am. J. Physiol. Cell Physiol.* **2005**, 288, 1.
- [36] E. A. Kabat. Structural Concepts in Immunology and Immunochemistry, New York: Holt, Rinehart and Winston, 2nd ed. **1976**, Chapter 9.
- [37] N. Shetty, E. Aarons, J. Andrews. Structure and functions of microbes. In *Infectious disease: pathogenesis, prevention, and case study*. N. Shetty, J. W. Tang, J. Andrews (Eds). John Wiley and Sons, **2009**, p.15.
- [38] D. Corbett, T. Hudson, I. S. Roberts. Bacterial polysaccharide capsules. In *Prokaryotic cell wall compounds*. H. König (Ed.). Springer, **2010**, p. 111.
- [39] D. M. Monack, A. Mueller, S. Falkow. Persistent bacterial infections: the interface of the pathogen and the host immune system. *Nature Rev. Microbiol.* **2004**, 2, 747.
- [40] C. F. Lin, S. C. Chiu, Y. L. Hsiao, S. W. Wan, H. Y. Lei, A. L. Shiau, H. S. Liu, T. M. Yeh, S. H. Chen, C. C. Liu, Y. S. Lin. Expression of cytokine, chemokine, and adhesion molecules during endothelial cell activation induced by antibodies against dengue virus nonstructural protein 1. *J. Immunol.* **2005**, 174, 395.
- [41] A. J. Pollard, K. P. Perrett, P. C. Beverley. Maintaining protection against invasive bacteria with protein-polysaccharide conjugate vaccines. *Nature Rev.* **2009**, 9, 213.
- [42] H. C. J. Gram. Über die isolirte färbung der schizomyceten in schnitt und trockenpräparaten. *Fortschritte der Medizin* **1884**, 2, 185.

- [43] E. Stackebrandt, R. G. E. Murray, H. G. Truper. Proteobacteria classis nov. A name for the phylogenetic taxon that includes the "purple bacteria and their relatives". *Int. J. Syst. Bacteriol.* **1988**, 38, 321.
- [44] A. M. Glauert, M. J. Thornley. The topography of the bacterial cell wall. *Annu. Rev. Microbiol.* **1969**, 23, 159.
- [45] T. J. Silhavy, D. Kahne, S. Walker. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, 1.
- [46] P. R. Murray. Medical Microbiology. **2002**, Mosby Inc.: <http://micro.digitalproteus.com/morphology2.php>
- [47] O. Westphal, O. Lüderitz, F. Bister. Über die Extraktion von Bakterien mit Phenol/Wasser. *Z. Naturforsch.* **1952**, 7B, 148.
- [48] E. Pupo, A. Aguila, H. Santana, J. F. Núñez, L. Castellanos-Serra, E. Hardy. Mice immunization with gel electrophoresis-micropurified bacterial lipopolysaccharides. *Electrophor.* **1999**, 20, 458.
- [49] D. Reisser, A. Pance, J. F. Jeannin. Mechanisms of the antitumoral effect of lipid A. *BioEssays* **2002**, 24, 284.
- [50] R. A. Bowden, A. Cloeckert, M. S. Zygmunt, G. Dubray. Outer-membrane protein- and rough lipopolysaccharide-specific monoclonal antibodies protect mice against *Brucella ovis*. *J. Med. Microbiol.* **1995**, 43, 344.
- [51] M. Fulop, P. Mastroeni, M. Green, R. W. Titball. Role of antibody to lipopolysaccharide in protection against low- and high-virulence strains of *Francisella tularensis*. *Vaccine* **2001**, 19, 4465.

- [52] G. Ada, D. Isaacs. Carbohydrate-protein conjugate vaccines. *Clin. Microbiol. Infect.* **2003**, 9, 79.
- [53] K. Landsteiner. In *The Specificity of Serological Reactions*. Cambridge: Harvard Univ. Press, **1945**, p. 156.
- [54] O. T. Avery, W. F. Goebel. Chemo-immunological studies on conjugated carbohydrate-proteins. II. Immunological specificity of synthetic sugar-protein antigens. *J. Exp. Med.* **1929**, 50, 533.
- [55] C. Soanes, A. Stevenson. Oxford Dictionary of English (2nd Ed.) (**2005**).
- [56] X. X. Gu, S. F. Rudy, C. Chu, L. McCullagh, H. N. Kim, J. Chen, J. Li, J. B. Robbins, C. Van Waes, J. F. Battey. Phase I study of a lipooligosaccharide-based conjugate vaccine against nontypeable *Haemophilus influenzae*. *Vaccine* **2003**, 21, 2107.
- [57] T. Wu, J. Malinverni, N. Ruiz, S. Kim, T. J. Silhavy, D. Kahne. Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell* **2005**, 121, 235.
- [58] S. Pal, E. M. Peterson, L. M. De la Maza. Vaccination with the *Chlamydia trachomatis* major outer membrane protein can elicit an immune response as protective as that resulting from inoculation with live bacteria. *Infect. Immun.* **2005**, 73, 8153.
- [59] A. Chernyak, S. Kondo, T. K. Wade, M. D. Meeks, P. M. Alzari, J. M. Fournier, R. K. Taylor, P. Kováč, W. F. Wade. Induction of protective immunity by synthetic *Vibrio cholerae* hexasaccharide derived from *V. cholerae* O1 Ogawa lipopolysaccharide bound to a protein carrier. *J. Infect. Dis.* **2002**, 185, 950.

- [60] R. Schneerson, O. Barrera, A. Sutton, J. B. Robbins. Preparation, characterization, and immunogenicity of *Haemophilus influenzae* type b polysaccharide-protein conjugates. *J. Exp. Med.* **1980**, *152*, 361.
- [61] J. B. Robbins, R. Schneerson, P. Anderson, D. H. Smith. The 1996 Albert Lasker Medical Research Awards. Prevention of systemic infections, especially meningitis, caused by *Haemophilus influenzae* type b. Impact on public health and implications for other polysaccharide-based vaccines. *J. Am. Med. Assoc.* **1996**, *276*, 1181.
- [62] X. Ma, R. Saxena, A. Chernyak, P. Kováč. Neoglycoconjugates from synthetic tetra- and hexasaccharides that mimic the terminus of the O-PS of *Vibrio cholerae* O:1, serotype Inaba. *Org. Biomol. Chem.* **2003**, *1*, 775.
- [63] J. Zhang, A. Yergey, J. Kowalak, P. Kováč. Studies towards neoglycoconjugates from the monosaccharide determinant of *Vibrio cholerae* O:1, serotype Ogawa using the diethyl squarate reagent. *Carbohydr. Res.* **1998**, *313*, 15.
- [64] M. D. Meeks, R. Saxena, X. Ma, T. K. Wade, R. K. Taylor, P. Kováč, W. F. Wade. Synthetic fragments of *Vibrio cholerae* O1 Inaba O-specific polysaccharide bound to a protein carrier are immunogenic in mice but do not induce protective antibodies. *Infect. Immun.* **2004**, *72*, 4090.
- [65] R. Saxena, A. Chernyak, A. Karavanov, P. Kováč. Conjugating low molecular mass carbohydrates to proteins 1. Monitoring the progress of conjugation. *Meth. Enzymol.* **2003**, *362*, 125.
- [66] R. Facklam, J. A. Elliott. Identification, classification, and clinical relevance of catalase-negative, gram-positive cocci, excluding the streptococci and enterococci. *Clin. Microbiol. Rev.* **1995**, *8*, 479.

- [67] C. R. Woese. Bacterial evolution. *Microbiol. Rev.* **1987**, *51*, 221.
- [68] J. Errington. Regulation of endospore formation in *Bacillus subtilis*. *Nature Rev. Microbiol.* **2003**, *1*, 117.
- [69] L. L. Matz, C. Beaman, P. Gerhardt. Chemical composition of exosporium from spores of *Bacillus cereus*. *J. Bacteriol.* **1970**, *101*, 196.
- [70] M. Garcia-Patrone, J. S. Tandecarz. A glycoprotein multimer from *Bacillus thuringiensis* sporangia: dissociation into subunits and sugar composition. *Mol. Cell Biochem.* **1995**, *145*, 29.
- [71] S. Charlton, A. J. G. Moir, L. Baillie, A. Moir. Characterization of the exosporium of *Bacillus cereus*. *J. Appl. Microbiol.* **1999**, *87*, 241.
- [72] P. Sylvestre, E. Couture-Tosi, M. Mock. A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol. Microbiol.* **2002**, *45*, 169.
- [73] D. B. Werz, P. H. Seeberger. Total synthesis of antigen *Bacillus anthracis* tetrasaccharide-Creation of an anthrax vaccine candidate. *Angewandte Chemie International Edition* **2005**, *44*, 6315.
- [74] M. Tajiri, S. Yoshida, Y. Wada. Differential analysis of site-specific glycans on plasma and cellular fibronectins: application of a hydrophilic affinity method for glycopeptide enrichment. *Glycobiol.* **2005**, *15*, 1332.
- [75] B. Imperiali, S. E. O'Connor. Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem Biol.* **1999**, *3*, 643.
- [76] T. Yue, B. B. Haab. Microarrays in glycoproteomics research. *Clin. Lab. Med.* **2009**, *29*, 15.

- [77] R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature*. **2003**, 422, 198.
- [78] W. Morelle, J. C. Michalski. Glycomics and mass spectrometry. *Curr Pharm Des*. **2005**, 11, 2615.
- [79] K. Dettmer, P. A. Aronov, B. D. Hammock. Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. **2007**, 26, 51.
- [80] S. J. Blanksby, T. W. Mitchell. Advances in Mass Spectrometry for Lipidomics. *Ann. Rev. Anal. Chem*. **2010**, 3, 433.
- [81] J. H. Banoub, R. P. Newton, E. Esmans, D. F. Ewing, G. Mackenzie. Recent developments in mass spectrometry for the characterization of nucleosides, nucleotides, oligonucleotides, and nucleic acids. *Chem. Rev*. **2005**, 105, 1869-1915.
- [82] C. S. Ho, C. W. K. Lam, M. H. M. Chan, R. C. K. Cheung, L. K. Law, L. C. W. Lit, K. F. Ng, M. W. M. Suen, H. L. Tai. Electrospray ionization mass spectrometry: principles and clinical applications. *Clin. Biochem. Rev*. **2003**, 24, 3.
- [83] N. B. Cech, C. G. Enke. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev*. **2001**, 20, 362.
- [84] L. Rayleigh. On the equilibrium of liquid conducting masses charged with electricity. *Philosophical Magazine* **1882**, 14, 184.
- [85] M. Abonnenc, L. Qiao, B. H. Liu, H. H. Girault. Electrochemical aspects of electrospray and laser desorption/ionization for mass spectrometry. *Annu. Rev. Anal. Chem*. **2010**, 3, 231.
- [86] E. De Hoffmann, V. Stroobant. In *Mass spectrometry: Principles and Applications*. Ed. John Wiley and Sons, **2007**, p.46.

- [87] M. Yamashita, J. B. Fenn. Electrospray ion source. Another variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88*, 4451.
- [88] C. K. Meng, M. Mann, J. B. Fenn. Of protons or proteins. "A Beam's a Beam for a that." (Burns OS). *Z. Phys. D. Atoms, Molecules and Clusters.* **1988**, *10*, 361.
- [89] H. Y. Kim, N. Jr. Salem. Application of thermospray high performance liquid chromatography/mass spectrometry for the determination of phospholipids and related compounds. *Anal. Chem.* **1987**, *59*, 722.
- [90] K. Strupat, M. Karas, F. Hillenkamp. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1991**, *72*, 89.
- [91] B. Guo, S. Wang, Y. Fan. Improving the performance of MALDI-TOF in oligonucleotide analysis using a new SDIFA technology. *Anal. Chem.*, **2000**, *72*, 5792.
- [92] R. G. Keck, J. B. Briggs, A. J. S. Jones. Oligosaccharide release and MALDI-TOF MS analysis of N-linked carbohydrate structures from glycoproteins. *Meth. Mol. Biol.* **2005**, *308*, 381.
- [93] B. Fuchs, J. Schiller. MALDI-TOF MS analysis of lipids from cells, tissues and body fluids. *Subcell. Biochem.* **2008**, *49*, 541.
- [94] A. Walch, S. Rauser, S. O. Deininger, H. Höfler. MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. *Histochem. Cell Biol.* **2008**, *130*, 421.
- [95] A. L. Gaertner. Increasing throughput and data quality for proteomics. In *Methods in Proteome and Protein Analysis*. R. M. Kamp, J. J. Calvete and T. Choli-Papadopoulou (Eds). Ed. Springer. **2002**, p. 383.

- [96] M. Karas, U. Baher, A. Ingedoh, E. Nordhoff, B. Stahl, K. Strupat, F. Hillenkamp. Principles and Applications of Matrix-assisted UV-laser Desorption/Ionization Mass Spectrometry. *Anal. Chem. Acta.* **1990**, 241, 175.
- [97] R. C. Beavis, B. T. Chait. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **1989**, 3, 432.
- [98] R. C. Beavis, B. T. Chait. α -Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, 27, 156.
- [99] *Tools of the Trade; Matrix Assisted Laser Desorption Ionization (MALDI)*. (n.d.). Retrieved May 5, 2009, from Magnet Lab; National High Magnetic Field Laboratory; FSU: http://www.magnet.fsu.edu/education/tutorials/tools/ionization_maldi.html
- [100] R. C. Beavis, B. T. Chait. Velocity distributions of intact high mass polypeptide molecule ions produced by matrix assisted laser desorption. *Chem. Phys. Lett.* **1991**, 181, 479.
- [101] R. Knochenmuss, R. Zenobi. MALDI ionization: the role of in-plume processes. *Chem. Rev.* **2003**, 103, 441.
- [102] K. L. Busch. Desorption ionization mass spectrometry. *J. Mass Spectrom.* **1995**, 30, 233.
- [103] V. V. Laiko, M. A. Baldwin, A. L. Burlingame. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **2000**, 72, 652.
- [104] V. V. Laiko, N. I. Taranenko, V. D. Berkout, B. D. Musselman, V. M. Doroshenko. Atmospheric pressure laser desorption/ionization on porous silicon. *Rapid Commun. Mass Spectrom.* **2002**, 16, 1737.

- [105] W. P. Peng, Y. C. Yang, M. W. Kang, Y. K. Tzeng, Z. Nie, H. C. Chang, W. Chang, C. H. Chen. Laser-induced acoustic desorption mass spectrometry of single bioparticles. *Angew. Chem. Int. Ed.* **2006**, *45*, 1423.
- [106] P. H. Dawson. Quadrupole mass spectrometry and its applications. Elsevier, Amsterdam, **1976** (reissued by AIP Press: Woodbury, NY, **1995**).
- [107] L. Yang, T. D. Mann, D. Little, N. Wu, R. P. Clement, P. J. Rudewicz. Evaluation of a four-channel multiplexed electrospray triple quadrupole mass spectrometer for the simultaneous validation of LC/MS/MS methods in four different preclinical matrices. *Anal. Chem.* **2001**, *73*, 1740.
- [108] O. Fiehn, J. Kopka, R. N. Trethewey, L. Willmitzer. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000**, *72*, 3573.
- [109] P. Schmitt-Kopplin, M. Frommberger. Capillary electrophoresis-mass spectrometry: 15 years of developments and applications. *Electrophor.* **2003**, *24*, 3837.
- [110] S. A. McLuckey, J. M. Wells. Mass Analysis at the Advent of the 21st Century. *Chem. Rev.* **2001**, *101*, 571.
- [111] A. El-Aneel. In Mass spectrometric analysis and fingerprint identification of natural lipopolysaccharide vaccine candidates and synthetic liposomal cholesterol neoglycolipids; Ph. D Thesis, Memorial University of Newfoundland, NL, **2006**, p 15.
- [112] W. E. Stephens. A pulsed mass spectrometer with time dispersion. *Phys. Rev.* **1946**, *69*, 691.
- [113] W. L. Wiley, I. H. McLaren. Time-of-flight mass spectrometer with improved resolution. *Rev. Sci. Instrum.* **1955**, *16*, 1150.

- [114] E. De Hoffmann, V. Stroobant. In *Mass spectrometry: Principles and Applications*. Wiley (Eds.), 3rd Edition, **2007**, p. 127.
- [115] A. Ingendoh, M. Karas, F. Hillenkamp, U. Giessmann. Factors affecting the mass resolution in matrix-assisted laser desorption-ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1994**, *131*, 345.
- [116] G. H. Herbert, R. A. W. Johnstone. In *Mass Spectrometry Basics*. Boca Raton: CRC Press LLC, **2003**, p. 189.
- [117] W. Pusch, M. Kostrzewa. Application of MALDI-TOF mass spectrometry in screening and diagnostic research. *Curr. Pharm. Des.* **2005**, *11*, 2577.
- [118] H. R. Morris, T. Paxton, M. Panico, R. McDowell, A. Dell. A Novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J. Prot. Chem.* **1997**, *16*, 469.
- [119] G. H. Herbert, R. A. W. Johnstone. In *Mass Spectrometry Basics*. Boca Raton: CRC Press LLC, **2003**, p. 172.
- [120] H. Oberacher, F. Pitterl. On the use of ESI-QqTOF-MS/MS for the comparative sequencing of nucleic acids. *Biopol.* **2009**, *91*, 401.
- [121] N. Tachon, F. Jahouh, M. Delmas, J. H. Banoub. Structural determination by atmospheric pressure photoionization tandem mass spectrometry of some compounds isolated from the SARA fractions obtained from bitumen. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2657.
- [122] A. Wattenberg, A. J. Organ, K. Schneider, R. Tyldesley, R. Bordoli, R. H. Bateman. Sequence dependent fragmentation of peptides generated by MALDI

quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 772.

[123] J. W. Flora, A. P. Null, D. C. Muddiman. Dual-micro-ESI source for precise mass determination on a quadrupole time-of-flight mass spectrometer for genomic and proteomic applications. *Anal. Bioanal. Chem.* **2002**, *373*, 538.

[124] P. Hufnagel, R. Rabus. Mass spectrometric identification of proteins in complex post-genomic projects. *J. Mol. Microbiol. Biotechnol.* **2006**, *11*, 53.

[125] I. V. Chernushevich, A. V. Loboda, B. A. Thomson. An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* **2001**, *36*, 849.

[126] A. Shevchenko, I. Chernushevich, W. Ens, K. G. Standing, B. Thomson, M. Wilm, M. Mann. Rapid *de novo* peptide sequencing by a combination of nanoelectrospray, isotopic labelling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015.

[127] R. E. March. Ion trap mass spectrometry. *Int. J. Mass Spectrom. Ion Processes.* **1992**, *118*, 71.

[128] R. G. Cooks, G. L. Glish, S. A. McLuckey, R. E. Kaiser. Ion trap mass spectrometry. *Chem. Eng. News* **1991**, *69*, 26.

[129] R. E. March, R. J. Hughes. In *Quadrupole Storage Mass Spectrometry*. Wiley, New York, **1989**.

[130] P. S. H. Wong, R. G. Cooks. Ion trap mass spectrometry. *Bioanal. Systems Inc.* **1997**, 16.

[131] D. J. Douglas, A. J. Frank, D. Mao. Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 1.

- [132] A. G. Marshall, C. L. Hendrickson, G. S. Jackson. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, 17, 1.
- [133] D. G. Schmid, P. Grosche, H. Bandel, G. Jung. FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry. *Biotechnol. Bioeng.* **2000/2001**, 71, 149.
- [134] X. Feng, M. M. Siegel. FTICR-MS applications for the structure determination of natural products. *Anal. Bioanal. Chem.* **2007**, 389, 1341.
- [135] S. W. Lee, S. J. Berger, S. Martinović, L. Paša-Tolić, G. A. Anderson, Y. Shen, R. Zhao, R. D. Smith. Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 5942.
- [136] J. J. Mulholland, S. M. Peterman. A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, 17, 168.
- [137] D. C. Muddiman, D. S. Wunschel, C. Liu, L. Paša-Tolić, K. F. Fox, A. Fox, G. A. Anderson, R. D. Smith. Characterization of PCR products from Bacilli using electrospray ionization FTICR mass spectrometry. *Anal. Chem.* **1996**, 68, 3705.
- [138] A. Makarov. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chem.* **2000**, 72, 1156.
- [139] J. H. Gross. In *Mass Spectrometry: A Textbook*. Springer, 2nd ed., **2011**, p. 191.
- [140] R. H. Perry, R. G. Cooks, R. J. Noll. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **2008**, 27, 661.
- [141] R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature* **2003**, 422, 198.

- [142] Z. M. Segu, L. A. Hammad, Y. Mechref. Rapid and efficient glycoprotein identification through microwave-assisted enzymatic digestion. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3461.
- [143] R. Breitling, A. R. Pitt, M. P. Barrett. Precision mapping of the metabolome. *Trends Biotechnol.* **2006**, *24*, 543.
- [144] H. Manduzio, E. Ezan, F. Fenaille. Evaluation of the LTQ-Orbitrap mass spectrometer for the analysis of polymerase chain reaction products. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3501.
- [145] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem.* **2006**, *78*, 2113.
- [146] E. De Hoffmann Tandem mass spectrometry: a Primer. *J. Mass Spectrom.* **1996**, *31*, 129.
- [147] F. W. McLafferty. In *Tandem mass spectrometry*. John Wiley and Sons, New York, **1983**.
- [148] W. J. Griffiths, A. P. Jonsson, S. Liu, D. K. Rai, Y. Wang. Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* **2001**, *355*, 545.
- [149] J. Li, S. M. Assmann. Mass spectrometry. An essential tool in proteome analysis. *Plant Physiol.* **2000**, *123*, 807.
- [150] J. V. Johnson, R. A. Yost, P. E. Kelley, D. C. Bradford. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *Anal. Chem.* **1990**, *62*, 2162.

- [151] R. A. Zubarev, N. L. Kelleher, F. W. McLafferty. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265.
- [152] M. I. Catalina, C. A. M. Koeleman, A. M. Deelder, M. Wührer. Electron transfer dissociation of *N*-glycopeptides: loss of the entire *N*-glycosylated asparagine side chain. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1053.
- [153] J. E. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528.
- [154] F. Sanger, A. R. Coulson. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **1975**, *94*, 441.
- [155] K. K. Murray. DNA sequencing by mass spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1203.
- [156] K. Boissinot, A. Huletsky, R. Peytavi, S. Turcotte, V. Veillette, M. Boissinot, F. J. Picard, E. A. Martel, M. G. Bergeron. Rapid exonuclease digestion of PCR-amplified targets for improved microarray hybridization. *Clin. Chem.* **2007**, *53*, 2020.
- [157] S. A. McLuckey, G. J. Van Berkel, G. L. Glish. Tandem mass spectrometry of small, multiply charged oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60.
- [158] C. M. Castleberry, M. Colette, L. P. Rodicio, P. Lenore, P. A. Limbach. Electrospray ionization mass spectrometry of oligonucleotides. In *Curr. Protoc. Nucleic Acid Chem.* S. L. Beaucage (Ed.) **2008**, Chapter 10, Unit 10.2.

- [159] A. Apffel, J. A. Chakel, S. Fischer, K. Lichtenwalter, W. S. Hancock. Analysis of oligonucleotide by HPLC-electrospray ionization mass spectrometry. *Anal. Chem.* **1997**, *69*, 1320.
- [160] J. Rozenski, Mongo Oligo Mass Calculator v2.06, **1999**:
<http://library.med.utah.edu/masspec/mongo.html>
- [161] J. Rozenski, J. A. McCloskey. SOS: a simple interactive program for *ab initio* oligonucleotide sequencing by mass spectrometry. *J. Am Soc. Mass Spectrom.* **2002**, *13*, 200.
- [162] A. I. Nesvizhskii. Protein identification by tandem mass spectrometry and sequence database searching. *Meth. Mol. Biol.* **2007**, *367*, 87.
- [163] M. Brosch, G. I. Saunders, A. Frankish, M. O. Collins, L. Yu, J. Wright, R. Verstraten, D. J. Adams, J. Harrow, J. S. Choudhary, T. Hubbard. Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. *Genome Res.* **2011**, *21*, 7567.
- [164] P. Alves, R. J. Arnold, M. V. Novotny, P. Radivojac, J. P. Reilly, H. Tang. Advancement in protein inference from shotgun proteomics using peptide detectability. *Pac. Symp. Biocomp.* **2007**, *12*, 409.
- [165] National Center for Biotechnology Information, U.S. National Library of Medicine:
<http://www.ncbi.nlm.nih.gov/>
- [166] J. R. Yates, J. K. Eng, A. L. McCormack, D. Schieltz. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **1995**, *67*, 1426.
- [167] Matrix Science Ltd, **2010**: <http://www.matrixscience.com/>

- [168] G. E. Reid, S. A. McLuckey. 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **2002**, 37, 663.
- [169] K. G. Standing. Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol.* **2003**, 13, 595.
- [170] P. Roepstorff, J. Fohlman. Letter to the editors. *Biol. Mass Spectrom.* **1984**, 11, 601.
- [171] R. S. Johnson, S. A. Martin, K. Biemann, J. T. Stults, J. T. Watson. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, 59, 2621.
- [172] B. Domon, C. E. Costello. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, 5, 397.
- [173] B. Domon, C. E. Costello. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochem.* **1988**, 27, 1534.
- [174] J. H. Banoub, A. El-Aneel, A. M. Cohen, P. Martin. Characterization of the O-4 phosphorylated and O-5 substituted Kdo reducing end group and sequencing of the core oligosaccharide of *Aeromonas Salmonicida* ssp *Salmonicida* lipopolysaccharide using tandem mass spectrometry. *Eur. J. Mass Spectrom. (Chichester, Eng)* **2004**, 10, 715.

CHAPTER 2: Differentiation and sequencing of three constitutional isobaric 18-mer DNA oligomers using low-energy collision tandem mass spectrometry

This chapter has been published: Farid Jahouh,¹ Mina Nashed,² Nicolas Joly,³ Patrick Martin³ and Joseph Banoub^{1,4*} *Int. J. Mass Spectrom.* **2011**, 304 (special issue: Nucleic Acids), 105.

¹Department of Chemistry, Memorial University of Newfoundland, St. John's, Newfoundland, A1B 3V6, Canada

²Department of Chemistry, Faculty of Science, Alexandria University Alexandria, Egypt

³Unité de Catalyse et de Chimie du Solide, site de l'Artois - UMR CNRS 8181, I.U.T. de Béthune, Département Chimie, 1230 rue de l'Université, BP819, 62408 Béthune cedex, France

⁴Fisheries and Oceans Canada, Science Branch, Environmental Sciences Division, Special Projects, P.O. Box 5667, St John's, Newfoundland, A1C 5X1, Canada

Abstract

In this manuscript we present the differentiation and sequencing of three constitutional isobaric 18-mer DNA oligomers, namely: **GATTCATAGCTACGAATC 1**, **AATTCGTAGCTACGAATC 2**, and **AATTCGTACCTACGAATG 3** using electrospray ionization-mass spectrometry with a hybrid QqTOF-MS/MS instrument.

The conventional single scan ESI-QqTOF-MS analyses of the DNA oligomers afforded the same series of deprotonated molecular ions. CID-MS/MS analyses were used to differentiate and to sequence the three DNA oligomers. Different CID-MS/MS analyses of the various deprotonated molecules: $[M - 8H]^{8-}$ at m/z 683.9881 and $[M - 9H]^{9-}$ at m/z 607.8775 were acquired for each 18-mer DNA oligomer. It was established that we could distinguish two different types of diagnostic product ions: a) the product ions formed by identical cleavage sites within each precursor anion, having the same isobaric masses and b) the product ions formed by identical cleavage sites, but having different masses. These specific diagnostic product ions facilitated the characterization of each isobaric 18-mer DNA oligomer. In addition, we also indicate that a series of product ions were common for at least two constitutional isobaric 18-mer oligomers. The Mongo Oligo Software was used exclusively to perform the characterization and the correct sequencing of each isobaric DNA oligomer.

2.1. Introduction

In modern molecular biotechnology, the analysis of DNA oligomers has become a topic of primary importance, since establishing the size, purity, and sequence of nucleic acids is a prerequisite to their use as molecular probes in medicinal science. The sequence of DNA encodes the necessary information for all living organisms to survive and reproduce. Therefore, knowledge of the DNA sequences is useful in biological research.^[1-3] In medicine, for example, DNA sequencing can be used to identify, diagnose and, potentially, to develop treatments for genetic diseases. Similarly, research into pathogens may lead to treatments for contagious diseases. Biotechnology is a burgeoning discipline, with the potential for many useful products and services.^[4]

Significant progress in the area of accurate mass determination, sequencing, and study of DNA oligomers and non-covalent interactions has been made possible by the use of novel mass spectrometric techniques.^[4-6] MALDI-TOF-MS has been explored widely for DNA sequencing.^[7-11] The Sanger dideoxy-procedure has been generally used for the sequencing of DNA-fragments. When compared to gel electrophoresis-based sequencing systems, mass spectrometry enables high resolution of the DNA-sequenced fragments, fast separation (on a microsecond time scale), and completely eliminates compressions that are associated with gel electrophoresis.^[8,9] In addition, the use of mass spectrometry for the analysis of the monomeric constituents of nucleic acids (nucleobases, nucleosides, nucleotides and derivatives) has been widely established in the literature for the last four decades.^[10-12] Before tandem mass spectrometry was used in conjunction with MALDI-MS, the gas phase fragmentation of the DNA anions were carried out with a TOF reflectron mass spectrometer using post-source decay.^[13-15]

Another successful method used for the sequencing of DNA is the ladder sequencing method.^[16] It consists of the chemical cleavage of the oligonucleotide with an exonuclease (phosphodiesterase for single strands of DNA and exonuclease III for double strands) digestion that sequentially removes nucleotides from the 3 - or the 5 - terminus of the oligomer.^[17-21] The exonuclease digestion was found to be both time and pH-dependent. Mass spectrometric analysis of the digestion products of the oligomer and the sequence of the unmodified or modified oligomer is resolved by the mass change after the loss of each nucleotide.^[17-21]

The ability of ESI-MS to form multiply charged ions in combination with improved mass range in novel state-of-the-art MS instruments has made possible the detection of increasingly larger oligonucleotide fragment ions. In reality, because DNA has an acidic phosphodiester backbone, oligonucleotides are usually analyzed by ESI-MS in the negative ion mode. Elucidation of the sequence of oligonucleotides by MS/MS of multiply charged DNA ions has been studied intensively over the past decade^[22] and ESI-MS and CID-MS/MS protocols have been established for the characterization of intact DNA oligomers that vary in size up to 70 residues.^[11,23-28]

Although CID-MS/MS analyses provide an excellent basis for detecting both the length and sequence variations of oligonucleotides,^[11] the presence of multiply charged ions can make interpretation of the spectra complicated. McLuckey *et al.* were the first to produce a nomenclature for the product ions obtained during the dissociation of multiply charged oligonucleotides (Figure 2.1(a)).^[29,30] It was discerned, that under moderate CID-MS/MS and MSⁿ conditions, further consecutive fragmentations occurred from previous charge-separation reactions. Thus, the presence of the complementary ions arising from

these decompositions greatly facilitated spectral interpretation.^[30,31] Furthermore, during the CID-MS/MS analysis of the studied oligonucleotides, the formation of internal product ions was also observed and these afforded valuable composition information on the sequence stretches. Needless to say, that the manual interpretation of the MS/MS analysis of oligonucleotides can be difficult, time-consuming, and prone to error because of the large number of product ions present. Consequently, efforts were made to simplify the process by relating the product ions obtained with the particular structure of a precursor oligonucleotide anion. A computer-based algorithm, to automatically derive sequence information from the MS/MS spectra for oligonucleotides of a completely unknown sequence, was developed by McCloskey and coworkers.^[32]

Rozenski and McCloskey developed a novel approach for the nearest-neighbor determination, based on the analysis of fragment ions of the nucleic acid formed in the ionization region of the mass spectrometer, along with the product ions obtained from MS/MS of the precursor oligonucleotide anion.^[33] The same authors wrote software for the *ab initio* determination of unknown oligonucleotide sequences up to the 20-mer level and was termed Simple Oligonucleotide Sequencer.^[34] However, this method cannot be extended to longer oligonucleotide sequences because missing fragments in the series of [a-B] or w ions would prevent the successful passage through the whole sequence.

The Mongo Oligo Mass Calculator v2.06 software was developed by Rozenski and collaborators and is available online.^[35] This software is used to determine the sequence of oligonucleotides provided by ESI-CID-MS/MS analysis of multi-charged ions and/or enzymatic digests. It is also capable of providing the molecular mass calculation of oligonucleotides and to predict the possible CID product ions $[a - B]$, $[d - H_2O]$, y and w . In addition, this software permits the determination of internal product ions (information about sequence stretches) that may be obtained for the analysis of a known oligonucleotide. Please note that the formation of internal product ions $[pd(A)pf]^-$ predicted by this program correspond to the following product ions represented as $[A_2:A_2]^-$, $[A_6:A_6]^-$, $[A_8:A_8]^-$, $[A_{12}:A_{12}]^-$, $[A_{15}:A_{15}]^-$ and $[A_{16}:A_{16}]^-$ for the oligonucleotide *I* (Figure 2.1(b)). The Mongo Oligo Program also calculates the possible mass of exo- and endonuclease digests ions.^[35]

The purpose of this work was to try to sequence, with a QqTOF-MS/MS hybrid instrument, an isobaric series of three self complementary 18-mer DNA oligomers which have the same base composition but different orders. The Mongo Oligo Program was used to obtain the CID-MS/MS correct sequencing. These 18-mer DNA oligomers are being used in our laboratory for the synthesis of model retrorsine derived-DNA adducts compounds used for the study. The retrorsine is a pro-carcinogen pyrrolizidine alkaloid, and its metabolites formed during liver microsomal activation are toxic and can bind DNA, causing cancer.^[36,37]

2.2. Material and methods

2.2.1. Sample preparation

The three constitutional isobaric 18-mer DNA oligomers used in our experiments were purchased from Oswel DNA Service Laboratory (Southampton, UK). These three self complementary 18-mer DNA oligomers were made up of 18 residues with the following sequences: GATTCATAGCTACGAATC 1, AATTCGTAGCTACGAATC 2, and AATTCGTACCTACGAATG 3, having the same base composition but different sequences.

2.2.2. Electrospray quadrupole orthogonal time-of-flight mass spectrometry

Conventional electrospray ionization-mass spectrometry for all the oligomers was acquired in the negative ion mode using an Applied Biosystems API-QSTAR XL quadrupole orthogonal time of-flight (QqTOF)-MS/MS hybrid tandem mass spectrometer (Applied Biosystems International-MDS Sciex, Foster City, California, USA). This instrument is capable of analyzing a mass range of m/z 50 to 10,000, with a resolution of 10,000 in the negative ion mode. ESI was performed with the turbo ionspray source operated at -4.2 kV. The ESI-MS were recorded with a cone voltage setting (declustering potential 1, DP1) of -50 V, a declustering potential 2 (DP2) of -10 V, a focusing potential (FP) of -255 V, a curtain gas of 20 Pascal (Pa), an ion source gas 1 (GS1) of 30 Pa, and an ion source gas 2 (GS2) of 30 Pa, which were generally kept constant for the analyses.

A solution of each oligonucleotide at a concentration of 1 mg mL⁻¹ was prepared by dissolving 1 mg of each sample in 1 mL of a mixture of acetonitrile/water (80/20) (v/v). Sample solution was then directly infused, with an integrated Harvard syringe

pump (Harvard Apparatus, Hollister, MA) at a rate of $5 \mu\text{L min}^{-1}$. One drop of diluted triethylamine was added to $500 \mu\text{L}$ of the oligonucleotide solutions with a concentration of 1 mg mL^{-1} , to increase the formation of the deprotonated molecular anions in the ion source of the mass spectrometer.

The TOF analyzer was calibrated for the mass range of the measured analytes using the following oligomer: TGATCA which was dissolved in a 1:1 mixture of acetonitrile (ACN or CH_3CN)-water (H_2O) to a final concentration of 1 mg mL^{-1} and checking for the exact masses of the $[\text{M} - \text{H}]^-$ at m/z 1789.3431 and $[\text{M} - 2\text{H}]^{2-}$ at m/z 894.1679, and using the TCA oligomer and checking for the exact mass of the $[\text{M} - \text{H}]^-$ at m/z 843.1870 and $[\text{M} - 2\text{H}]^{2-}$ at m/z 421.0899.

2.2.3. Low-energy collision CID-MS/MS analyses

Product ion scan experiments were recorded with the same instrument. The product ion spectra were obtained from fragmentation in the QqTOF-MS/MS hybrid instrument. In the product ion scan mode, the first quadrupole (Q1) selected the corresponding precursor ion. The precursor ion was fragmented into product ions by collision with nitrogen in the radio-frequency (RF), linear acceleration pulsar high pressure (LINAC) equipped, quadrupole collision cell. Nitrogen collision gas was added to the enclosed chamber of the quadrupole (Q2) to give an indicated pressure of 1×10^{-5} torr for collisional activation of the sample ions. The collision energy (CE) and the CID gas conditions ensured that the precursor ion remained abundant. The product ions were scanned and sorted in the orthogonal mass resolving time-of-flight analyzer which measured the occurrence of particular product ions, previously formed in the Q2. We have found the collision energies for the CID-MS/MS of deprotonated molecules of these

oligomers did not exceed 40 eV while the declustering potential was kept at the same value (DP1 = -50 V) as that of the electrospray full scan ESI-MS experiments.

2.3. Results and discussion

Ni, Chang and Oberacher *et al.* have studied the sequence verification of different length oligonucleotides using ESI-QTOF-MS.^[38,39] The results were also compared to those obtained with an ESI triple quadrupole MS and an ESI Ion Trap MS. They found that the resolving power of the QTOF and its sensitivity allowed better interpretation of the spectra and also easier sequencing of the oligonucleotides. It is important to mention that the resolving power of the QqTOF-MS instrument can only reach 20,000 full width at half maximum (FWHM)^[39,40] which is a moderate resolution of lower orders of magnitude when compared to high resolution-mass spectrometry such as Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS or FT-MS) or the recent LTQ Orbitrap XL Hybrid FT which can reach up to 1,000,000 FWHM.^[41,42] In this rationale, the ESI-QqTOF-MS and CID-MS/MS analyses of this series of isobaric 18-mer oligonucleotides were performed using a QqTOF-MS/MS hybrid instrument. This latter instrument allowed us to identify and separate the isotopic deprotonated molecules without ambiguity.

In addition, we have opted to use reliable software (Mongo Oligo Mass Calculator) for the identification and characterization of the CID-product ions obtained from the product ion scan analyses of the well resolved, diagnostic isotopic deprotonated molecules, selected of our isobaric 18-mer oligomers. The Mongo Oligo Mass Calculator uses the sequence of a know oligonucleotide as input for the calculations. It can be used for calculating masses of oligonucleotides, electrospray series, fragment product ions

obtained by collision-induced dissociation (CID), and fragments of enzymatic digests by endonuclease and exonuclease.^[35] The method consists on comparing the product ion spectra masses generated by CID of multiply charged oligonucleotide ions to the predicted m/z values, employing established fragmentation pathways obtained from a known reference sequence. The program can also calculate the masses of modified residues. An accompanying tool is Oligo Composition Calculator which finds oligonucleotide composition when the user inputs a mass. Mongo Oligo Mass Calculator is by far the most comprehensive program for the analysis of genetic mass spectrometric data with multiple functionalities.^[35,43]

2.3.1. ESI-QqTOF-MS analyses

The ESI-QqTOF-MS (negative ionization mode) of the three isobaric 18-*mer* DNA oligomers showed identical series of multi-charged deprotonated molecular anions (Table 2.1, Figure A.1). Seeing that the identical molecular weight of the isobaric 18-*mer* DNA oligomers is high, the intensity of the monoisotopic mass of the deprotonated molecules tends to be lower than the nominal and average masses. As expected, ESI-QqTOF-MS analysis of the three constitutional isobaric 18-*mer* DNA oligomers gave the equivalent series of highly charged deprotonated molecular anions with the same m/z values. It is also well known that the DNA molecular anions, when highly charged, are extremely easy to fragment.^[30,31,33-35,43] In addition, the three analyzed isobaric 18-*mer* oligomers, formed in the gas phase the same series of sodiated adduct anions. The theoretical monoisotopic m/z values of the deprotonated molecules and their Δ deviations (in ppm) are presented in Table 2.1.

Table 2.1. Molecular and fragment ions obtained during the conventional ESI-QqTOF-MS analysis of the oligomers **1-3**.

Ion	Calculated <i>m/z</i>	Deviation for 1 (ppm)	Deviation for 2 (ppm)	Deviation for 3 (ppm)
[M - 11H] ¹¹⁺	497.1712	14	15	17
[M + Na - 12H] ¹¹⁺	499.1695	9	11	-18
[M - 10H] ¹⁰⁺	546.9890	11	-13	18
[M + Na - 11H] ¹⁰⁺	549.1860	2	0	-4
[M + K - 11H] ¹⁰⁺	550.7830	-1	4	-
[M + 2Na - 12H] ¹⁰⁺	551.3840	1	8	-34
[M - 9H] ⁹⁺	607.8775	-3	-4	29
[M + Na - 10H] ⁹⁺	610.3200	-10	-5	39
[M + K - 10H] ⁹⁺	612.0948	-16	-11	-
[M + 2Na - 11H] ⁹⁺	612.7624	-18	-18	13
[M - 8H] ⁸⁺	683.9881	-7	-3	1
[M + Na - 9H] ⁸⁺	686.7340	-6	-8	-3
[M + K - 9H] ⁸⁺	688.731	-13	-13	-
[M + 2Na - 10H] ⁸⁺	689.4836	-15	-7	2
[M - 7H] ⁷⁺	781.8446	-7	-9	-25
[M + Na - 8H] ⁷⁺	784.9849	-9	-9	-44
[M + K - 8H] ⁷⁺	787.2669	-12	-6	-
[M + 2Na - 9H] ⁷⁺	788.1252	-12	-3	-48
[M - 6H] ⁶⁺	912.3199	0	0	-8
[M + Na - 7H] ⁶⁺	915.9836	2	-1	-59
[M + K - 7H] ⁶⁺	918.6459	-6	-4	-
[M + 2Na - 8H] ⁶⁺	919.6472	-5	-2	-24
[M + 3Na - 9H] ⁶⁺	923.3109	-22	-15	-51
[M - 5H] ⁵⁺	1094.9853	-5	-1	-9
[M + Na - 6H] ⁵⁺	1099.3817	-7	7	-38
[M + 2Na - 7H] ⁵⁺	1103.7781	-3	7	-55
[M + 3Na - 8H] ⁵⁺	1108.1745	-32	-10	-2

2.3.2. Low-energy collision CID-MS/MS analyses

In general, during all the tandem analyses presented herein, we performed on the extracted isotopic precursor ions. The CID-MS/MS experiments of the deprotonated molecules $[M - 8H]^{8-}$ at m/z 683.9881 and $[M - 9H]^{9-}$ at m/z 607.8775, were acquired for each isobaric 18-mer DNA oligomer. Initial attempts to generate product-ion scans from the multi-charged oligonucleotide anion precursors relied on maximizing the abundances of the product ions of high masses to obtain as much information as possible on each different sequence. Thus, the best CE was chosen for each molecular anion, which was about 30 eV. As already reported in the literature, we have found that when the analyzed anion was highly charged, it needed less CE to fragment.^[44]

2.3.3. CID-MS/MS analysis of the $[M - 8H]^{8-}$ deprotonated molecules at m/z 683.9881 extracted from the isobaric 18-mer DNA oligomers *I-3*

We have noticed that each individual product ion scan created a series of different diagnostic ions (observed only in one oligonucleotide), as well as common product ions (Figure 2.2 for oligomer *I*). The specific product ions obtained by CID-MS/MS analysis of the $[M - 8H]^{8-}$ ion at m/z 683.9881 of the 18-mer DNA oligomers *I-3* are shown in Figure 2.3 and Tables A.1-A.3. It is important to point out that the examination of each diagnostic product ion indicated that their formations were not localized within the same sequential region of the constitutional 18-mer oligomers. Therefore, the formations of these diagnostic product ions are unique and constitute a fingerprint of the analyzed isobaric 18-mer oligomers. On the other hand, the CID-MS/MS fragmentations of the deprotonated molecules, yielded identical mini-sequences, occurring at the same cleavage positions; however, the product ions formed had different charged states. This latter

observation may be due to the existence of different deprotonation sites of the respective analyzed oligomers. In addition, it is well known that the final CID-MS/MS fragmentation pattern obtained can be influenced by the nucleobases present in each of the analyzed 18-mer DNA oligomers.^[12] It is logical to expect that isobaric oligonucleotides with different sequences basically have different physical-chemical properties, which can be the reason for the differences observed during their gas-phase fragmentations.

Additionally, we observed the formation complementary product ions, which helped in the product ion identification. McLuckey *et al.* found that when a base is eliminated from the deprotonated molecule $[M - nH]^{n-}$ ($n > 1$), it produces the anion $[M - nH - B]^{(n-1)-}$ (B = lost base), which fragments further to afford the complementary pair of product ions w_a^{x-} and $[a_m - B]^{k-}$ in which $x, k > 1$ and $(x + k) = (n - 1)$ (Figure 2.1(a)).^[30,31] Another fragmentation pathway leading to the formation of complementary product ions involves the loss of a neutral base (A, T, C or G) to produce the following anion: $[M - nH - BH]^{n-}$ (BH = neutral base loss) which also fragments, leading to the formation of the following complementary product ions: w_a^{x-} and $[a_m - B]^{k-}$ ($x, k > 1$ and $(x + k) = n$).

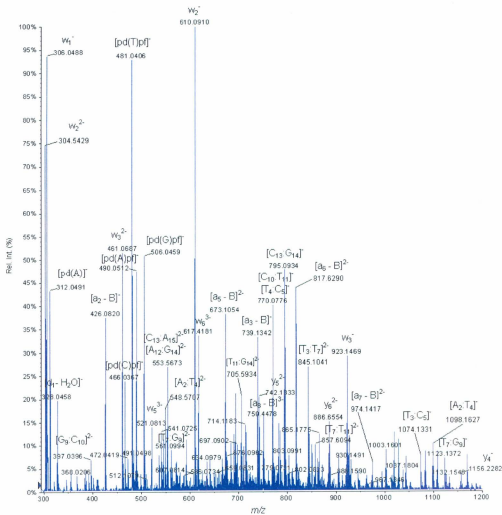


Figure 2.2. Low-energy CID tandem mass spectra of the $[M - 8H]^{8-}$ anion at m/z 683.9881 of the oligomer *I*.

Thus, in the CID-MS/MS analysis of the multi-charged precursor $[M - 8H]^{8-}$ ion, the loss of a single base (A, T, C or G) can produce the hepta-charged product anion: $[M - 8H - B]^{7-}$ which probably fragments spontaneously and therefore is not detected. However, this $[M - 8H - B]^{7-}$ product ion is able to afford the following complementary product ions. For the oligomer **1** (Figure 2.6(a)), we identified complementary product ions: $[a_{16} - B(A)]^{6-}$ at m/z 787.9272 and w_2^{-} at m/z 610.0910, and $[a_{12} - B(A)]^{4+}$ at m/z 871.3856 and w_6^{3+} at m/z 617.4181. However, no complementary product ions were observed for the constitutional isobaric DNA oligomer **2** (Figures 2.3(b) and 2.6(b)). For the DNA oligomer **3**, we identified the complementary product ions $[a_8 - B(A)]^{3-}$ at m/z 750.4395 and w_{10}^{4+} at m/z 771.6299; $[a_{12} - B(A)]^{4+}$ at m/z 861.3868 and w_6^{3+} at m/z 630.7709 and $[a_{16} - B(A)]^{6-}$ at m/z 781.2867 and w_2^{-} at m/z 650.0968 (Figures 2.3(c) and 2.6(c)). Furthermore, a neutral base loss during the CID-MS/MS analysis of the precursor ion $[M - 8H]^{8-}$ is able to form the octa-charged product ion $[M - 8H - B]^{8-}$ (not detected in the spectra), which is an intermediate ion for the production of the following complementary product ions: $[a_{16} - B(A)]^{6-}$ at m/z 787.9272 and w_2^{2-} at m/z 304.5429 and $[a_{12} - B(A)]^{5-}$ at m/z 696.9452 and w_6^{3+} at m/z 617.4181 for the oligonucleotide **1** (Figures 2.3(a) and 2.6(a)). Also, we were not able to identify any complementary product ions for the oligonucleotide **2** (Figures 2.3(b) and 2.6(b)). Finally, the complementary product ions $[a_{16} - B(A)]^{6-}$ at m/z 781.2867 and w_2^{2-} at m/z 324.5523; $[a_{15} - B(A)]^{6-}$ at m/z 729.0918 and w_3^{2-} at m/z 481.0689; $[a_{12} - B(A)]^{4+}$ at m/z 861.3868 and w_6^{4+} at m/z 472.8377 and $[a_8 - B(A)]^{4+}$ at m/z 562.5952 and w_{10}^{4+} at m/z 771.6299 were detected for the oligonucleotide **3** (Figures 2.3(c) and 2.6(c)).

The CID-MS/MS analyses confirm without doubt, the known fact, that these three 18-mer DNA oligomers, while being isobaric, have indeed different sequences.

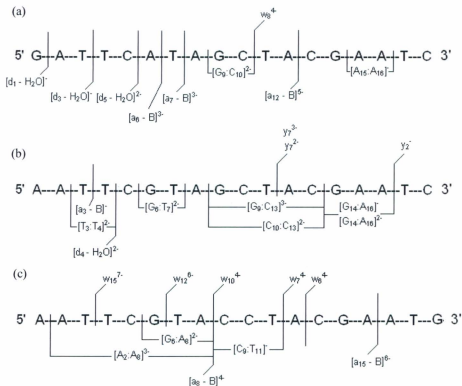


Figure 2.3. Different specific product ions obtained during the CID-MS/MS analysis of the $[M - 8H]^{8-}$ ion at m/z 683.9881, for the following (a) oligomer 1, (b) oligomer 2 and (c) oligomer 3.

2.3.4. CID-MS/MS analysis of the $[M - 9H]^{9-}$ deprotonated molecules at m/z

607.8775 extracted from the isobaric 18-mer DNA oligomers 1-3

The CID-MS/MS analysis of the $[M - 9H]^{9-}$ ion at m/z 607.8775 for the oligonucleotide *1* (Figure 2.4), afforded different diagnostic product ions (Tables A1-A3). The cleavage sites within each analyzed respective precursor ion are shown in Figure 2.5. As expected, we have identified a series of diagnostic complementary product ions that were tentatively assigned as being created from the non-detected $[M - 9H - B]^{8-}$ intermediate product ion. Accordingly, we have identified for oligomer *1* the following complementary product ions: $[a_{12} - B(A)]^{5-}$ at m/z 696.8982 and w_6^{3-} at m/z 617.4415; $[a_{12} - B(A)]^{4+}$ at m/z 871.3750 and w_6^{4+} at m/z 462.8181; and $[a_{16} - B(A)]^{6-}$ at m/z 787.9570 and w_2^{2-} at m/z 304.5428 (Figures 2.5(a) and 2.7(a)). It is interesting to note that for the CID-MS/MS analyses of the $[M - 9H]^{9-}$ ion at m/z 607.8775 extracted from the 18-mer oligomers *2* and *3* we did not observe the formation of any complementary product ions (Figures 2.5(b) and 2.7(b), and 2.5(c) and 2.7(c)). Unfortunately, we were not able to identify the pairs of complementary product ions arising from the fragmentation of the intermediary nona-charged product ion $[M - 9H - B]^{9-}$ (not detected), which should be produced during the CID-MS/MS analysis of the precursor ion $[M - 9H]^{9-}$ via the loss of a neutral base. It is evident that the CID-MS/MS of the precursor $[M - 9H]^{9-}$ ion at m/z 607.8775 affords less complementary product ions than the analysis of the $[M - 8H]^{8-}$ precursor ion.

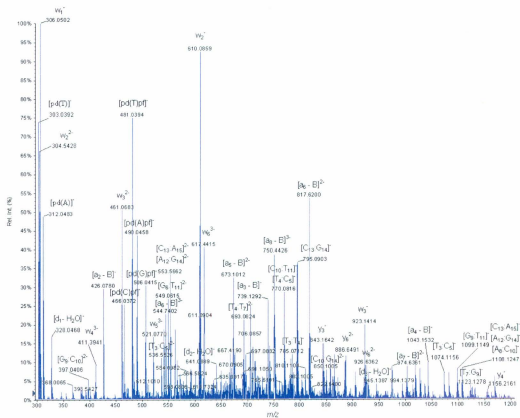


Figure 2.4. Low-energy CID tandem mass spectra of the $[M - 9H]^{9-}$ anion at m/z 607.8775 of the oligomer 1.

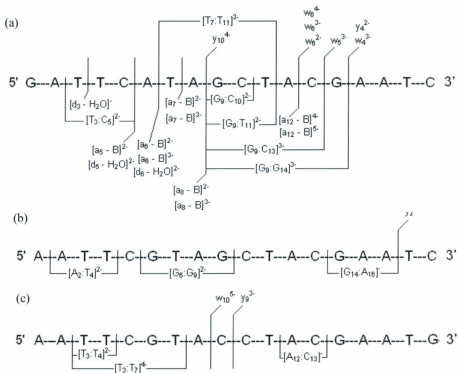


Figure 2.5. Different specific product ions obtained during the CID-MS/MS analysis of the $[M - 9H]^3$ ion at m/z 607.8775, for the following (a) oligomer 1, (b) oligomer 2 and (c) oligomer 3.

2.3.5. Common product ions present in the product ion scans of the oligomers 1-3

During each single CID-MS/MS analysis conducted with the different series of precursor ions selected from the oligomers 1-3, we have observed a common series of either identical or similar product ions shown in Tables A.4 and A.5. We have tried to classify them into two types: a) the product ions formed by identical cleavage sites within each precursor anion, having the same masses which corresponded to product ions with the same bases composition and b) the product ions formed by identical cleavage sites, however having different masses, corresponding to product ions with a different bases composition. Accordingly, depending on the cleavage site, some product ions may have identical masses.

Thus, the CID-MS/MS of the precursor ions $[M - 8H]^{8-}$ at m/z 683.9881 and $[M - 9H]^{9-}$ at m/z 607.8775 extracted from GATTCATAGCTACGAATC 1 (Figures 2.6(a) and 2.7(a), Tables A.2 and A.3) and AATTCGTAGCTACGAATC 2 (Figures 2.6(b) and 2.7(b), Tables A.2 and A.3) have the following common product ions with the same masses (the common sequence from each termini of the two oligonucleotides is displayed in bold): w_n and y_n ($n \leq 12$), $[a_m - B]$ ($m \geq 7$) and $[d_p - H_2O]$ ($p \geq 6$). Indeed, during the CID-MS/MS analysis of the $[M - 8H]^{8-}$ at m/z 683.9881 the common product ions, with the same masses of these two oligomers, are: w_2^{2-} at m/z 304.5429, w_1^- at m/z 306.0488, w_3^{2-} at m/z 461.0687, w_5^{3-} at m/z 521.0813, y_4^{2-} at m/z 577.6099, w_2^- at m/z 610.0910, w_6^{3-} at m/z 617.4181, y_5^{2-} at m/z 742.1333, $[a_8 - B]^{3-}$ at m/z 750.4478, w_5^{2-} at m/z 782.1231, y_3^- at m/z 843.1699, y_6^{2-} at m/z 886.6554, w_3^- at m/z 923.1469, w_6^{2-} at m/z 926.6393 and $[a_7 - B]^{2-}$ at m/z 974.1417.

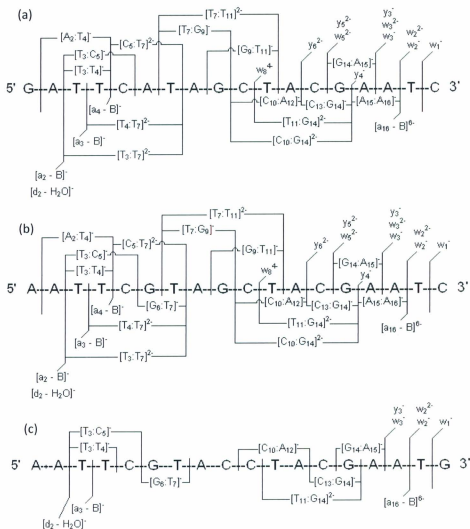


Figure 2.7. Common product ions (isobars and non isobars) to at least two oligonucleotides obtained during the CID-MS/MS analysis of the $[M - 9H]^{-9}$ ion at m/z 607.8775, for the following (a) oligomer 1, (b) oligomer 2 and (c) oligomer 3.

The product ion scan of the $[M - 9H]^{9-}$ at m/z 607.8775 gives us the following common product ions having the same masses of the two oligonucleotides: w_2^{2-} at m/z 304.5428, w_1^{-} at m/z 306.0502, w_3^{2-} at m/z 461.0683, w_2^{-} at m/z 610.0859, w_8^{4-} at m/z 617.0860, y_5^{2-} at m/z 742.1295, w_5^{2-} at m/z 782.1193, y_3^{-} at m/z 843.1642, y_6^{2-} at m/z 886.6485, w_3^{-} at m/z 923.1414 and y_4^{-} at m/z 1156.2161.

The comparison of the CID-MS/MS of the precursor ion $[M - 8H]^{8-}$ at m/z 683.9881 extracted from the **GATTCATAGCTACGAATC 1** and the **AATTCGTACCTACGAATG 3** (differences in the bases location between the two oligomers are displayed in bold) (Figures 2.6(a) and 2.6(c)) revealed the formation of one common product ion $[a_8 - B]^{1-}$ at m/z 750.4478 having the same mass (Table A.4). However, it is worth noting that this product ion has the same base composition, but with a different sequence for the two oligonucleotides. This is not surprising, as in the CID-MS/MS analyses of the deprotonated molecules $[M - 8H]^{8-}$ at m/z 683.9881 and also $[M - 9H]^{9-}$ at m/z 607.8775 obtained from oligomers **1** and **3**, the extracted precursor ions have different termini and therefore cannot form the common product ions w_n , y_n , $[a_m - B]$ and $[d_p - H_2O]$ with the same sequence. Please note that, in the CID-MS/MS of the precursor ion $[M - 9H]^{9-}$ extracted from **GATTCATAGCTACGAATC 1** and the **AATTCGTACCTACGAATG 3** no common product ion having the same mass were observed.

Finally, the comparison between the CID-MS/MS of the precursor ions $[M - 8H]^{8-}$ at m/z 683.9881 (Figure 2.6, Table A.4) and $[M - 9H]^{9-}$ at m/z 607.8775 (Figure 2.7, Table A.5) extracted from **AATTCGTAGCTACGAATC 2** and **AATTCGTACCTACGAATG 3** produced the following common product ions having

the same masses: w_n and y_n ($n \geq 10$), $[a_m - B]$ ($m \leq 9$) and $[d_p - H_2O]$ ($p \leq 8$). Indeed, during the CID-MS/MS of the precursor ions $[M - 8H]^{8-}$ at m/z 683.9881 we observed the following common product ions having the same mass: $[a_2 - B]^-$ at m/z 410.0818, $[a_5 - B]^{2-}$ at m/z 665.1079, w_{17}^{7-} at m/z 748.5498 and $[a_8 - B]^{3-}$ at m/z 750.4530. From these observations, it was construed that, although these common product ions contain the same bases, they do not necessarily have the same sequence.

We also observed the formation of common internal product ions with the same masses which correspond to the same sequence. For example, the internal product ion $[T_{11}:G_{14}]^{2-}$ at m/z 705.5935 of the CID-MS/MS analysis of the $[M - 8H]^{8-}$ molecular ion at m/z 683.9881 (and $[M - 9H]^{9-}$ at m/z 607.8775) correspond to the following sequence $[pd(TACG)pf]^{2-}$ for the oligomers **1-3** (Figures 2.6 and 2.7, Tables A.4 and A.5). We can also say that the product ions w_n and y_n ($n \leq 12$) correspond to part of a same sequence for the **GATTCATAGCTACGAATC 1** and **AATTCGTAGCTACGAATC 2** oligomers (the common sequence from each termini of the two oligonucleotides is displayed in bold), as well as the product ions $[a_m - B]$ ($m \leq 9$) and $[d_p - H_2O]$ ($p \leq 8$) which have the same sequence for the **AATTCGTAGCTACGAATC 2** and **AATTCGTACCTACGAATG 3**. However in the case of the CID-MS/MS of $[M - 8H]^{8-}$ at m/z 683.9881 and $[M - 9H]^{9-}$ at m/z 607.8775 extracted from **GATTCATAGCTACGAATC 1** and **AATTCGTACCTACGAATG 3** oligomers do not have common product ions. This is explained by the fact that these two oligonucleotides **1** and **3** have different bases at their 3'- and 5'- termini. In addition, these constitutional isobaric 18-mer DNA oligomers can break down from both their 3'- and 5'- termini.

Furthermore, we also have observed the formation of internal product ions, having the same masses, but different sequences. For example, the internal product ion $[C_5:T_7]^{2+}$ at m/z 541.0725 correspond to the ion $[pd(CAT)pf]^{2+}$ for the oligonucleotide *1*. Unfortunately, this product ion at m/z 541.0725, can also correspond to the following ions: $[pd(TAC)pf]^{2+}$, $[pd(ATC)pf]^{2+}$ and $[pd(CAT)pf]^{2+}$ for the oligonucleotide *1*; $[pd(TAC)pf]^{2+}$ and $[pd(ATC)pf]^{2+}$ for the oligonucleotide *2* and $[pd(TAC)pf]^{2+}$ and $[pd(CTA)pf]^{2+}$ for the oligonucleotide *3* (Tables A.4 and A.5). In addition, the internal product ion $[C_5:T_7]^{2+}$ represents the sequence $[pd(CGT)pf]^{2+}$ at m/z 549.0629 for the oligomers *2* and *3*. We can conclude from this example, that the internal product ions can give limited information about the sequence stretches of the oligonucleotides.

In the CID-MS/MS analyses of the precursor ions $[M - 8H]^{8+}$ at m/z 683.9881 and $[M - 9H]^{9+}$ at m/z 607.8775 extracted from the 18-mer DNA oligomers *1-3*, we noticed the presence of the product ions arising from the same specific site cleavages but which have different masses. Thus, the GATTCATAGCTACGAATC *1* and AATTCGTAGCTACGAATC *2* give the following common product ions with different masses: w_n and y_n ($n \geq 13$), $[a_m - B]$ ($2 \leq m \leq 6$) and $[d_p - H_2O]$ ($p \leq 5$). The GATTCATAGCTACGAATC *1* and AATTCGTACCTACGAATG *3* have the following common product ions which differ in their masses: w_n and y_n ($n \leq 9$ and $13 \leq n \leq 17$), $[a_m - B]$ ($2 \leq m \leq 6$ and $m \geq 10$) and $[d_p - H_2O]$ ($p \leq 5$ and $9 \leq p \leq 17$).

Finally, for the pair AATTCGTAGCTACGAATC *2* and AATTCGTACCTACGAATG *3*, we identified the common product ions, which have not the same masses: w_n and y_n ($n \leq 9$), $[a_m - B]$ ($m \geq 10$) and $[d_p - H_2O]$ ($9 \leq p \leq 17$).

2.4. Conclusion

The constitutional isobaric 18-mer DNA oligomers GATTCATAGCTACGAATC, AATTCGTAGCTACGAATC and AATTCGTACCTACGAATG 1-3 (Molecular Mass = 5479.9560 Da) were analyzed by negative electrospray mass spectrometry. The resolution of the QqTOF mass spectrometer allowed us to observe the isotopic distribution of each ion, which is helpful for the identification of the molecular ions and the fragments ions.

The conventional ESI-QqTOF-MS of these 18-mer DNA oligomers exhibited identical series of multi-charged deprotonated molecular ions. Low-energy collision-induced dissociation tandem mass spectrometric analysis of the multi-charged oligonucleotide anions $[M - 8H]^{8-}$ at m/z 683.9881 and $[M - 9H]^{9-}$ at m/z 607.8775 provided characteristic and distinct fingerprint patterns which permitted discrimination amongst the individual oligomers and allowed complete bi-directional sequence verification. On the other hand, the *ab initio* sequencing was not possible for the 18-mer oligonucleotides because complicated CID-MS/MS spectra derived from highly charged precursor ions were obtained.

Finally, we would like to indicate that the Mongo Oligo Mass Calculator v2.06 software, developed by Rozenski, is an excellent program which facilitates the characterization and sequencing of DNA oligomers and helped to decipher the complex CID-MS/MS analyses obtained in this work.

Appendix A. Supplementary data

Supplementary are data associated with this chapter (page 231)

References

- [1] S. Pourshahian, P.A. Limbach. Application of fractional mass for the identification of peptide-oligonucleotide cross-links by mass spectrometry. *J. Mass Spectrom.* **2008**, *43*, 1081.
- [2] A. Durairaj, P. A. Limbach. Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection. *Anal. Chim. Acta* **2008**, *612*, 173.
- [3] A. E. Egger, C. G. Hartinger, H. Ben Hamidane, Y. O. Tsybin, B. K. Keppler, P. J. Dyson. High resolution mass spectrometry for studying the interactions of cisplatin with oligonucleotides. *Inorg. Chem.* **2008**, *47*, 10626.
- [4] P. F. Crain. Mass spectrometric techniques in nucleic acid research. *Mass Spectrom. Rev.* **1990**, *9*, 505.
- [5] J. H. Banoub, J. Miller-Banoub, F. Jahouh, N. Joly, P. Martin. In *Mass Spectrometry of Nucleosides and Nucleic Acids*, CRC Press LLC, 1st edition **2009**, p. 1.
- [6] M. C. Fitzgerald, L. Zhu, M. M. Smith. The analysis of mock DNA sequencing reactions using matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 895.
- [7] M. S. Roskey, P. Juhasz, I. P. Smirnov, E. J. Takach, S. A. Martin, L. A. Haff. DNA sequencing by delayed extraction-matrix assisted laser desorption/ionization time of flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 4724.
- [8] S. Mouradian, D. R. Rank, L. M. Smith. Analyzing sequencing reactions from bacteriophage M13 by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1475.

- [9] F. Kirpekar, E. Nordhoff, L. K. Larsen, K. Kristiansen, P. Roepstorff, F. Hillenkamp. DNA sequence analysis by MALDI mass spectrometry. *Nucleic Acids Res.* **1998**, 26, 2554.
- [10] D. Fu, K. Tang, A. Braun, D. Reuter, B. Darnhofer-Demar, D. P. Little, M. J. O'Donnell, C. R. Cantor, H. Koster. Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. *Nature Biotechnol.* **1998**, 16, 381.
- [11] T. R. Covey, R. F. Bonner, B. I. Shushan, J. Henion. The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, 2, 249.
- [12] S. A. McLuckey, S. Habibi-Goudarzi. Ion trap tandem mass spectrometry applied to small multiply charged oligonucleotides with a modified base. *J. Am. Soc. Mass Spectrom.* **1994**, 5, 740.
- [13] B. Spengler. Post-source decay analysis in matrix-assisted laser desorption/ionization mass spectrometry of biomolecules. *J. Mass Spectrom.* **1997**, 32, 1019.
- [14] H. M. Liebich, S. Müller-Hagedorn, F. Klaus, K. Meziane, K. R. Kim, A. Frickenschmidt, B. J. Kammerer. Chromatographic, capillary electrophoretic and matrix-assisted laser desorption ionization time-of-flight mass spectrometry analysis of urinary modified nucleosides as tumor markers. *J. Chromatogr. A* **2005**, 1071, 271.
- [15] E. Nordhoff, F. Kirpekar, P. Roepstorff. Mass spectrometry of nucleic acids. Mass spectrometry of nucleic acids. *Mass Spectrom. Rev.* **1996**, 15, 67.

- [16] G. Talbo, M. Mann. Aspects of the sequencing of carbohydrates and oligonucleotides by matrix-assisted laser desorption/ionization post-source decay. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 100.
- [17] U. Pielele, W. Zurcher, M. Schar, H. E. Moser. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. *Nucleic Acids Res.* **1993**, *21*, 3191.
- [18] C. M. Bentzley, M. V. Johnson, B. S. Larsen, S. Gutteridge. Oligonucleotide sequence and composition determined by matrix-assisted laser desorption/ionization. *Anal. Chem.* **1996**, *68*, 2141.
- [19] I. P. Smirnov, M. T. Roskey, P. Juhasz, E. J. Takach, S. A. Martin, L. A. Haff. Sequencing oligonucleotides by exonuclease digestion and delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Anal. Biochem.* **1996**, *238*, 19.
- [20] P. A. Limbach, P. F. Crain, J. A. McCloskey. Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **1994**, *22*, 2183.
- [21] H. Wu, C. Chan, H. Aboleneen. Sequencing regular and labeled oligonucleotides using enzymatic digestion and ionspray mass spectrometry. *Anal. Biochem.* **1998**, *263*, 129.
- [22] J. H. Banoub, R. P. Newton, E. Esmans, D. F. Ewing, G. MacKenzie. Recent developments in mass spectrometry for the characterization of nucleosides, nucleotides, oligonucleotides, and nucleic acids. *Chem. Rev.* **2005**, *105*, 1869.

- [23] A. P. Null, L. M. Benson, D. C. Muddiman. Enzymatic strategies for the characterization of nucleic acids by electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2699.
- [24] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64.
- [25] R. D. Smith, J. A. Loo, C. G. Edmonds, C. J. Barinaga, H. R. Udseth. New developments in biochemical mass spectrometry: electrospray ionization. *Anal. Chem.* **1990**, *62*, 882.
- [26] J. T. Stults, J. C. Marsters. Improved electrospray ionization of synthetic oligodeoxynucleotides. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 359.
- [27] R. D. Smith, J. A. Loo, R. R. Ogorzalek Loo, M. Busman, H. R. Udseth. Principles and practice of electrospray ionization-mass spectrometry for large polypeptides and proteins. *Mass Spectrom. Rev.* **1991**, *10*, 359.
- [28] P. A. Limbach, P. F. Crain, J. A. McCloskey. Characterization of oligonucleotides and nucleic acids by mass spectrometry. *Curr. Opin. Biotechnol.* **1995**, *6*, 96.
- [29] S. A. McLuckey, G. J. Van Berkel, G. L. Glish. Tandem mass spectrometry of small, multiply charged oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60.
- [30] S. A. McLuckey, S. Habibi-Goudarzi. Decompositions of multiply charged oligonucleotide anions. Decompositions of multiply charged oligonucleotide anions. *J. Am. Chem. Soc.* **1993**, *115*, 12085.
- [31] S. A. McLuckey, G. Vaidyanathan, S. Habibi-Goudarzi. Charged vs. neutral nucleobase loss from multiply charged oligonucleotide anions. *J. Mass Spectrom.* **1995**, *30*, 1222.

- [32] J. Ni, S. C. Pomerantz, J. Rozenski, Y. Zhang, J. A. McCloskey. Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry. *Anal. Chem.* **1996**, *68*, 1989.
- [33] J. Rozenski, J. A. McCloskey. Determination of nearest neighbors in nucleic acids by mass spectrometry. *Anal. Chem.* **1999**, *71*, 1454.
- [34] J. Rozenski, J. A. McCloskey. SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 200.
- [35] <http://library.med.utah.edu/masspec/mongo.htm>
- [36] P. S. Chu, M. W. Lame, H. J. Segall. In vivo metabolism of retrorsine and retrorsine-N-oxide. *Arch. Toxicol.* **1993**, *67*, 39.
- [37] A. R. Mattocks, I. N. White. Toxic effects and pyrrolic metabolites in the liver of young rats given the pyrrolizidine alkaloid retrorsine. *Chem-Biol. Interactions.* **1973**, *6*, 297.
- [38] J. Ni, K. Chan. Sequence verification of oligonucleotides by electrospray quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1600.
- [39] H. Oberacher, H. Niederstätter, W. Parson. Characterization of synthetic nucleic acids by electrospray ionization quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* **2005**, *40*, 932.
- [40] I. V. Chernushevich, A. V. Loboda, B. A. Thomson. An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* **2001**, *36*, 849.

- [41] S. Shi. Comparison and interconversion of the two most common frequency-to-mass calibration functions for Fourier transform ion cyclotron resonance mass spectrometry. *Int. J. Mass Spectrom.* **2000**, *195*, 591.
- [42] Makarov, E. Denisov, O. Lange, S. Horning. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977.
- [43] Q. Liao, C. Shen, P. Vouros. GenoMass - a computer software for automated identification of oligonucleotide DNA adducts from LC-MS analysis of DNA digests. *J. Mass Spectrom.* **2009**, *44*, 549.
- [44] J. A. Loo, H. R. Udseth, R. D. Smith. Collisional effects on the charge distribution of ions from large molecules, formed by electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 207.

CHAPTER 3: Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry

This chapter has been published: Farid Jahouh,¹ Rina Saksena,² Donatella Aiello,³ Anna Napoli,³ Giovanni Sindona,³ Pavol Kováč² and Joseph H. Banoub^{1,4*} *J. Mass Spectrom.* **2010**, *45*, 1148.

¹Department of Chemistry, Memorial University of Newfoundland, Saint John's, NL, A1B 3V6, Canada

²NIDDK, LBC, National Institutes of Health, Bethesda, MD 20892-0815, USA

³Department of Chemistry, University of Calabria, Arcavacata di Rende (CS), Italy

⁴Department of Fisheries and Oceans Canada, Science Branch, Special Projects, Saint John's, NL, Canada, A1C 5X1

Abstract

We present the MALDI-TOF/TOF-MS analyses of various hapten-BSA neoglycoconjugates obtained by squaric acid chemistry coupling of the spacer-equipped, terminal monosaccharide of the *O*-specific polysaccharide (*O*-PS) of *Vibrio cholerae* O1, serotype Ogawa, to BSA. These analyses allowed not only to calculate the molecular masses of the hapten-BSA neoglycoconjugates with different hapten:BSA ratios (4.3:1, 6.6:1 and 13.2:1) but, more importantly, also to localize the covalent linkages (conjugation sites) between the hapten and the carrier protein. Determination of the site of glycation was based on comparison of the MALDI-TOF/TOF-MS analysis of the peptides resulting from the digestion of BSA with similar data resulting from the digestion of BSA-glycoconjugates, followed by sequencing by MALDI-TOF/TOF-MS/MS of the glycated peptides. The product ion scans of the protonated molecules were carried out with a MALDI-TOF/TOF-MS/MS tandem mass spectrometer equipped with a high CE cell. The high-energy CID spectra afforded product-ions formed by fragmentation of the carbohydrate hapten and amino acid sequences conjugated with fragments of the carbohydrate hapten. We were able to identify three conjugation sites on lysine residues (Lys 235, Lys 437 and Lys 455). It was shown that these lysine residues are very reactive and bind lysine specific reagents. We presume that these Lys residues belong to those which are considered to be sterically more accessible on the surface of the 3-dimensional structure. The identification of the γ -series product ions was very useful for the sequencing of various peptides. The series of α - and β -product-ions confirmed the sequence of the conjugated peptides.

3.1. Introduction

During the last century, LPS have been extensively studied because of their importance as essential components of outer membranes in gram-negative bacteria.^[1] The LPS macromolecule consists of an internal core oligosaccharide attached to the lipid A, and an external core oligosaccharide which is glycosylated by a polysaccharide, also called the *O*-specific polysaccharide (*O*-PS). The latter is composed of oligosaccharide repeating units whose structure and composition differ for various genera and bacterial serotypes.^[2,3] Although, the LPSs are known to activate the complement and macrophages, they can also induce endotoxic shock. The toxic properties of the LPSs reside largely in the lipid A, whereas the characteristic immunological properties reside in the *O*-PS.

It has been established that *O*-specific polysaccharides are the essential virulence factors and protective antigens of pathogenic gram-negative bacteria.^[4] Many LPS-derived neoglycoconjugates, which included *O*-specific antigens, synthetic oligosaccharide antigens and delipidated LPS, have been used as protective vaccines in numerous therapeutic studies.^[11-9] It has been demonstrated that protein conjugates of *O*-PS of several enteropathogenic bacteria are safe and immunogenic in humans. They can elicit high titers of serum antibodies, which can be boosted by subsequent vaccinations, and offer protection against the homologous bacteria. In addition, it has been established that the efficacy of the neoglycoconjugate vaccines depend mainly on the structural (architectonic) variables. These include, but are not limited to, the size of the antigenic saccharide, the average number of saccharide chains per conjugate molecule, the distance

between the saccharide and the protein, the site of attachment to the protein, and the nature of the carrier protein.^[10-12]

Cholera is a disease caused by the gram-negative bacteria *Vibrio cholerae*, which causes severe bacterial infection of the intestine. The infection occurs in endemic areas or anywhere by ingesting contaminated food or liquids.^[13] The *Vibrio cholerae* strain belongs to the *Vibrionaceae* family of bacteria which is divided into some 200 serogroups.^[14] Kováč's group, has been involved in developing a conjugate vaccine for cholera from synthetic fragments of the O-PS of *Vibrio cholerae* O:1 conjugated by single point attachment to protein carriers. They have prepared well-defined neoglycoconjugates using spacer-equipped synthetic oligosaccharides attached to squaric acid monoesters. The oligosaccharides used mimic fragments of the O-specific polysaccharide of the two main strains of *Vibrio cholerae* serogroup O1, serotypes Ogawa (Figure 3.1(a)) and Inaba. The O-PS of both strains is composed of a chain of about 15-20 α -(1 \rightarrow 2)-linked monomers of 4-amino-4,6-dideoxy-D-mannopyranosyl residues (D-perosamine) in which the amino group is acetylated with 3-deoxy-L-glycero-tetronic acid.^[15,16] The O-PSs of the two strains differ in that only in the Ogawa strain is the O-2 of the upstream^[17] terminal perosamine residue methylated.^[18-20]

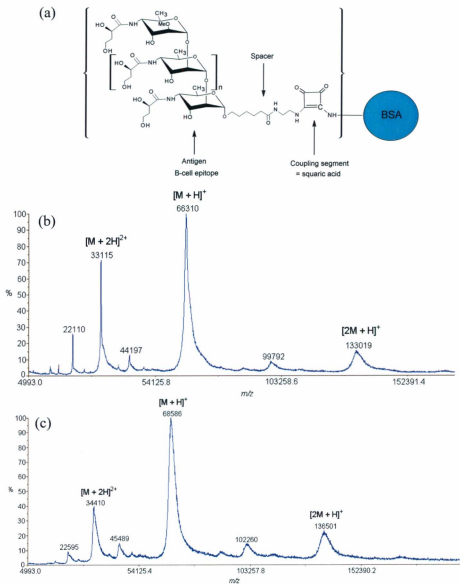


Figure 3.1. (a) Schematic representation of the general structure of carbohydrate-spacer-squaric acid-protein constructs from oligosaccharide fragments of the *O*-polysaccharide of the *Vibrio cholerae* O1, serotype Ogawa and (b) MALDI-TOF/TOF-MS analysis of native BSA, and (c) hapten-BSA neoglycoconjugates with a hapten:protein ratio of 4.3:1.

It is well known that the ϵ -amino groups of lysine residues in proteins are very reactive and, therefore, amenable to covalent attachments of haptens. The effect of attaching a hydrophilic carbohydrate hapten to such amino groups substantially alters the physicochemical and functional immunological properties of the carrier protein. There are 60 lysine residues in the BSA which are distributed in three homologous α -helical domains (I, II and III), and theoretically, all can be derivatized.^[21,22] Unfortunately, methods suitable for determination of the composition of neoglycoconjugates reveal only the carbohydrate:protein ratios but cannot localize the exact lysine residues where the haptens are attached. In addition, the carbohydrate:protein ratios in neoglycoconjugates are usually determined by colorimetric methods and not by MS. Regrettably, the reliability of calibration curves required by colorimetry have always been questionable, especially when they are to be used to analyze mixtures where oligosaccharides that mimic the structure of rare or unstable sugar-containing bacterial polysaccharides are in mixture with variable amounts of protein carriers.

Currently, several methods have been developed for the MS analysis of *N*-glycoproteins.^[23-25] ESI and MALDI techniques have been used to evaluate the distribution of glycation within the glycated hemoglobin molecule and for the determination of advanced glycation end-products on proteins associated with aging and various diseases.^[26-28] Recently, the characterization of the glycation sites of bovine insulin was determined by MALDI-MS/MS.^[29] However, reports on detailed structural elucidation of the synthesized neoglycoconjugate vaccines by mass spectrometry has been quite meager. Little work has been done with regard to localizing the lysine residues where the carbohydrate haptens are attached in protein carriers. Recently, the

carbohydrate:protein ratio (loading) in synthetic neoglycoconjugates has been measured objectively by MALDI-TOF/TOF-MS and SELDI-TOF-MS.^[30-34] However, neither these sophisticated techniques can locate the carbohydrate-bearing lysine residues.

With the task outlined above in mind, we have used a conventional approach whereby the digestion of glycoprotein with a protease was followed by mass spectral analysis of the resulting glycopeptides. Accordingly, we have compared the MALDI-TOF/TOF-MS analysis of the intact glycoconjugates composed of the simple antigenic glycosyl hapten the 1-[(2-aminoethylamido)carbonylpentyl-4-*N*-methylamino-4,6-dideoxy- α -D-mannopyranosyl]-2-methoxycyclobutene-3,4-dione conjugated to BSA to that of the BSA alone. In addition, we also compared the MALDI-TOF/TOF-MS of the peptides resulting from the digestion of the BSA to those resulting from the digestion of the hapten-BSA neoglycoconjugates. The sequencing of the resulting peptides was performed by high-energy collision dissociation CID-TOF/TOF-MS/MS using the selected protonated molecules formed from the released glycated peptides.

3.2. Materials and Methods

3.2.1. Preparation of the hapten-BSA neoglycoconjugates

The hapten-BSA conjugates were synthesized as described previously.^[34] Briefly, the 4-(3-deoxy-L-*glycero*)-2-*O*-methyl- α -D-perosamine, which corresponds to the upstream terminal moiety of the *O*-PS of *Vibrio cholerae* serogroup O1 serotype Ogawa antigen, was first prepared in the form of methyl 6-hydroxyhexanoyl α -glycoside, whose aglycone formed a spacer. The reaction of the spacer-equipped sugar with 1,2-diaminoethane converted the methyl ester group in the spacer into the corresponding

amino amide, and the subsequent reaction with diethyl squarate resulted in formation of squarate monoester. Finally, the conjugation with the BSA was carried out in commercial borate buffer (pH 9.0) at the initial molar carbohydrate:protein ratio of 100:1. The process of conjugation was monitored by SELDI-TOF-MS, and portions of the conjugation mixtures were withdrawn at various reaction times,^[34] to give the corresponding BSA conjugates. Isolation and purification of the hapten-BSA neoglycoconjugates was achieved by ultrafiltration, using centrifugal devices, to remove low molecular mass materials present (buffer salts and the unchanged carbohydrate derivative). According to the SELDI-TOF-MS analysis performed by Kováč's group, the carbohydrate:protein ratios were found to be 4.4:1, 7.1:1 and 13.4:1.^[34] When re-measured during this work the carbohydrate:protein ratios found were 4.3:1, 6.6:1 and 13.2:1.

3.2.2. Digestion

The hapten-BSA conjugate was reduced with 10 mM dithiothreitol (DTT) and alkylated with 100 mM iodoacetamide, followed by tryptic digestion using a solution of 20 ng/mL trypsin in NH_4HCO_3 (50 mM) and keeping the samples at 37 °C (water bath) overnight. The sample was then dried under vacuum and reconstituted in 20 μL of 1% acetic acid. Before mass spectral analysis, an aliquot of each sample (10 μL) was purified using ZipTip C18 (Millipore, Bedford, MA).

3.2.3 MALDI-TOF-MS and MALDI-TOF/TOF-MS/MS analyses

Mass spectral analysis were carried out with a 4700 Proteomics analyzer with TOF-TOF optics (Applied Biosystems Foster City, CA) and a 200-Hz frequency-tripled Nd:YAG laser. CHCA was used as matrix for the analysis of BSA, hapten-BSA conjugate, and the peptides resulting from their digestions, with an average of 5000 to 8000 laser shots per spectra. The MS/MS analyses were achieved with air as collision gas at collision energy of 1 kV, which corresponded to the difference between the accelerating potential (8 kV) and the floating cell (7 kV). The internal calibration of the digested glycoproteins and BSA was performed in the range 600-3200 using as standards peptides known to be present in the BSA.

3.3. Results

3.3.1. MALDI-TOF/TOF-MS analysis of BSA and the hapten-BSA neoglycoconjugates

To determine the number of haptens conjugated to the BSA, the BSA used as the starting carrier and the carbohydrate hapten-BSA neoglycoconjugates were analysed by single stage MS using MALDI-TOF/TOF-MS. Figure 3.1 shows the spectra of the starting BSA (b) and the synthetic neoglycoconjugate with 4.3:1 hapten:BSA ratio (c). We also recorded the MALDI-TOF/TOF-MS of the synthetic neoglycoconjugate with 6.6:1 and 13.2:1 hapten:BSA ratios, which gave the expected spectra (Figure B.1, Appendix B). The molecular masses of these synthetic neoglycoconjugates are listed in Table 3.1. The squaric acid chemistry of conjugation covalently links the carbohydrate squaric acid ethyl ester to the ϵ -amino function of the lysine residues in BSA with the

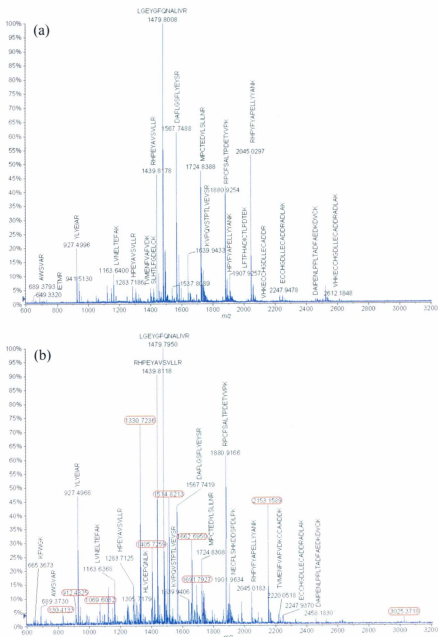
concomitant loss of EtOH. Consequently, knowing that the molecular mass of the monosaccharide-squaric acid portion was 513.23 Da, it became possible to calculate and verify the hapten:BSA ratio in the formed neoglycoconjugates. As expected, neoglycoconjugates resulting from longer reaction time^[34] showed higher ratio of conjugated haptens to the BSA carrier. Accordingly, the conjugate formed after the reaction time of 45 min, 1.5 h and 6 h showed carbohydrate:protein ratios of 4.3:1, 6.6:1 and 13.2:1, respectively.

Table 3.1. Molecular masses and calculated hapten:BSA ratios of BSA and hapten-BSA conjugates.

	$[M + 2H]^{2+}$	$[M + H]^+$	$[2M + H]^+$	Remeasured Hapten:BSA ratio
Standard BSA	33,115	66,310	133,019	-
Hapten:BSA 4:1	34,410	68,586	136,501	4.3:1
Hapten:BSA 7:1	34,659	70,125	139,462	6.6:1
Hapten:BSA 13:1	36,498	73,413	145,870	13.2:1

To identify the specific glycation sites in the formed neoglycoconjugates, the samples were enzymatically digested with trypsin, as described in the Materials and Methods section. Figure 3.2 displays the MALDI-TOF/TOF-MS analysis of peptides resulting from the digestion of the BSA (a) and of those from the hapten-BSA neoglycoconjugate with an initial ratio of 4.3:1 (b), 6.6:1 (c) and 13.2:1 (d). Identities of the diagnostically significant peptides and glycated peptides are shown in Tables 3.2 and 3.3. We have assigned only the peaks which appeared in the MALDI spectra of the digested hapten-BSA neoglycoconjugates (circled in the MALDI spectra, Figure 3.2). Accordingly, the following peaks corresponded to the glycated peptides at m/z 830.41, m/z 912.48, m/z 1069.60, m/z 1330.72, m/z 1405.73, m/z 1514.82, m/z 1662.70, m/z 1691.79, m/z 2153.16 and m/z 3025.37 (Figure 3.2). Unquestionably, these peaks were

assigned to the glycated peptides, in which the resulting molecular masses corresponded to the exact addition of the glycan mass increment of 513.23 Da for the targeted digested tryptic peptides. It is interesting to note that for the glycated peptide ions at m/z 1330.72 and 1514.82, the peptide portions were different from any peptides formed during the conventional digestion of the BSA carrier protein. However, the unique peak at m/z 2153.16 appears to be associated to an observed unglycated peptide at m/z 1639.94 formed during the conventional enzymatic digestion of BSA. The formation of peptides with m/z 1330.72, 1514.82 and 2153.16 is listed in the Table 3.3. These glycated peptides had a mass shift of 513.23 Da (characteristic of the glycation with one hapten residue) when compared with the m/z values of peptides (theoretical values). Note that the starting BSA is a commercial product, which has already gone through some purification/precipitation process. Therefore the glycated peptides were assigned as SLGK*VGTR (T63-T64, 452-459) at m/z 817.49, ALK*AWSVAR (T34-T35, 233-241) at m/z 1001.59 and K*VPQVSTPTLVEVSR (T61-T62, 437-451) at m/z 1639.94. The conjugation/glycation site that points to the lysine residue in the carrier protein is denoted by an asterisk. We were able to detect some other unidentified peptides (not formed during the conventional BSA digestion) and these were suspected to be glycated. However, their CID-MS/MS analyses did not give any useful information permitting their sequencing.



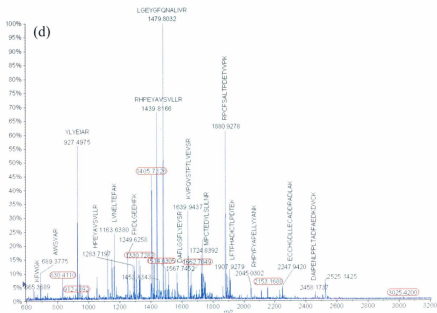
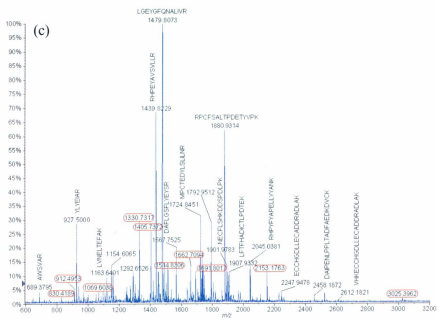


Figure 3.2. (continued)

Table 3.2. MALDI-TOF/TOF-MS peptide mapping of the digested BSA (after recalibration).

Peptide #	Location	Observed m/z	Calculated m/z	Accuracy ppm	Missed cleavage	Peptide sequence
T46-T48	310-340	3511.6710	3511.6725	0	2	SHCIAEVEKDAIPENLPPLTADFAEDKDVCK
T40-T42	264-285	2612.1848	2612.1655	-7	2	VHKECCHGDLLECCADRADLAK
T47-T48	319-340	2458.1887	2458.1812	-3	1	DAIPENLPPLTADFAEDKDVCK
T41-T42	267-285	2247.9478	2247.9433	-2	1	ECCHGDLLECCADRADLAK
T40-T41	264-280	2113.8802	2113.8853	2	1	VHKECCHGDLLECCADDR
T23-T24	168-183	2045.0297	2045.0285	-1	0	RHPYFYAPELLYYANK
T73	529-544	1907.9257	1907.9213	-2	0	LFTFHADICTLPDTEK
T24	169-183	1888.9312	1888.9265	-2	0	HPYFYAPELLYYANK
T71	508-523	1880.9254	1880.9216	-2	0	RPCFSALTPDETYVVPK
T66	469-482	1724.8388	1724.8351	-2	0	MPCTEDYLSLIINR
T61-T62	437-451	1639.9433	1639.9383	-3	0	KVPQVSTPTLVEVSR
T50	347-359	1567.7488	1567.7433	-4	0	DAFLGSFLYEYSR
T59	421-433	1479.8008	1479.7960	-3	0	LGEYGFQNALIVR
T51-T52	360-371	1439.8178	1439.8123	-4	1	RHPEYAVSVLLR
T12	89-100	1419.7009	1419.6942	-5	0	SLHTLFGDELCK
T79	569-580	1399.6932	1399.6931	0	0	TVMENFVAFVDK
T57	402-412	1305.7236	1305.7167	-5	0	HLVDEPONLIK
T52	361-371	1283.7186	1283.7112	-6	0	HPEYAVSVLLR
T7-T8	35-44	1249.6273	1249.6217	-4	1	FKDLGEEHFK
T10	66-75	1163.6400	1163.6312	-8	0	LVNELTEFAK
T22	161-167	927.4996	927.4940	-6	0	YLYEIAR
T35	236-241	689.3793	689.3735	-8	0	AWSVAR
T27	205-209	649.3320	649.3343	4	0	IETMR
T13	101-105	545.3421	545.3411	-2	0	VASLR

Table 3.3. MALDI-TOF/TOF-MS peptide mapping of the digested neoglycoconjugate with a hapten:protein ratio of 4.3:1, after recalibration (glycated peptides bolted).

Peptide #	Location	Observed m/z	Calculated m/z	Accuracy ppm	Missed cleavage	Sequence
T46-T48	310-340	3511.6660	3511.6725	2	2	SHCIAEVEKDAIPENLPPLTADFAEDKDVCK
T47-T48	319-340	2458.1830	2458.1812	-1	1	DAIPENLPPLTADFAEDKDVCK
T41-T42	267-285	2247.9370	2247.9433	3	2	ECCHGDLLECADDRADLAK
T78-T79	569-587	2220.0518	2219.9775	-33	1	TVMENFVAFVDKCCAADDK
T61-T62	437-451	2153.1589	2153.1700	5	0	K*VPQVSTPTLVEVSR
T23-T24	168-183	2045.0183	2045.0285	5	0	RHPYFYAPELLYYANK
T16-T17	123-138	1901.9634	1901.8703	-49	2	NECFLSHKDDSPDLPK
T71	508-523	1880.9166	1880.9216	3	0	RPCFSALTPDETYVPK
T66	469-482	1724.8308	1724.8351	2	0	MPCTEDYLSILNR
T61-T62	437-451	1639.9406	1639.9383	-1	1	KVPQVSTPTLVEVSR
T50	347-359	1567.7419	1567.7433	1	0	DAFLGSFLYEYSR
T34-T35	233-241	1514.8213	1514.8213	0	1	ALK*AWSVAR
T59	421-433	1479.7950	1479.7960	1	0	LGEYGFQNALIVR
T51-T52	360-371	1439.8118	1439.8123	0	1	RHPEYAVSVLLR
T12	89-100	1419.7020	1419.6942	-5	0	SLHTLFGDELCK
T63-T64	452-459	1330.7236	1330.7213	-2	1	SLGK*VGTR
T57	402-412	1305.7179	1305.7167	-1	0	HLVDEPQNLIK
T52	361-371	1283.7125	1283.7112	-1	0	HPEYAVSVLLR
T7-T8	35-44	1249.6237	1249.6217	-2	1	FKDLGEEHFK
T10	66-75	1163.6368	1163.6312	-5	0	LVNELTEFAK
T22	161-167	927.4966	927.4940	-3	0	YLYEIAR
T35	236-241	689.3730	689.3735	1	0	AWSVAR
T20-T21	156-160	665.3673	665.3775	15	1	KFWGK
T13	101-105	545.3493	545.3411	-15	0	VASLR

3.3.2. High-energy CID-MS/MS analyses of the digested BSA and hapten-BSA conjugates

To locate the glycation sites, the peptides were analysed by CID-TOF/TOF-MS/MS (Figure 3.3, Tables B.1-B.3). Analysis of the CID-MS/MS spectrum of the glycated peptide with m/z 1330.72 (Figure 3.3(a), Table B.1) allowed us to confirm the sequence of the peptide and to localize the carbohydrate hapten adduct, which resulted to the attachment to the Lys 455 residue. The sequencing of this glycated peptide was determined by identifying the y-series of product-ions: SLGK*VGTR (452-459). We also noted that the carbohydrate hapten undergoes fragmentation during the MS/MS analysis. Accordingly, simple sugar fragments were also present in the CID spectra (Figure 3.4), as evidenced by the presence of the following product-ions: $[C_6H_8NO_3]^+$ at m/z 142.07, $[C_{11}H_{18}NO_5]^+$ at m/z 244.13, $[C_{11}H_{20}NO_6]^+$ at m/z 262.14 and $[C_{15}H_{30}NO_7]^+$ at m/z 336.21. In addition, we noted the presence of the precursor molecular ion that lost a hapten fragment and the following product-ions: $[M + H - C_{11}H_{19}NO_6]^+$ at m/z 1069.59 and $[M + H - C_{11}H_{19}NO_6 - NH_3]^+$ at m/z 1052.57, where $C_{11}H_{19}NO_6$ represents a product ion fragment of the carbohydrate-spacer moiety. The CID spectrum also contained fragments of the a-, b- and z-series formed from product ions representing portions of the carbohydrate-spacer peptide fragments: $[a_4 - C_{15}H_{28}NO_7]^{++}$ at m/z 537.33, $[b_4 - C_{11}H_{19}NO_6]^+$ at m/z 638.39, $[b_5 - C_{11}H_{19}NO_6]^+$ at m/z 737.46 and $[z_6 - C_{11}H_{19}NO_6]^+$ at m/z 852.48. The secondary cleavage of the carbohydrate-spacer portion gives rise to additional specific product ions of the carbohydrate-spacer moiety. These also include the presence of a radical product ion, as for example the $[a_4 - C_{15}H_{28}NO_7]^{++}$ ion at m/z 537.33, which allowed the unambiguous characterization of the carbohydrate moiety as

the radical product ion obtained contains the carbohydrate unit, in addition to a small part of the spacer. Therefore, the presence of this radical product ion and the other product ions with carbohydrate-spacer moieties represents diagnostic product ion signatures. Their formation is the result of the labile fragmentation of the carbohydrate-spacer portion due to its instability to the high-energy collision, which takes place during the MS/MS experiment. The presence of these product ions complicates the CID spectra, making them more difficult to interpret. It is important to note that the peptide sequencing described here was performed without the help of a data sequencing based program.

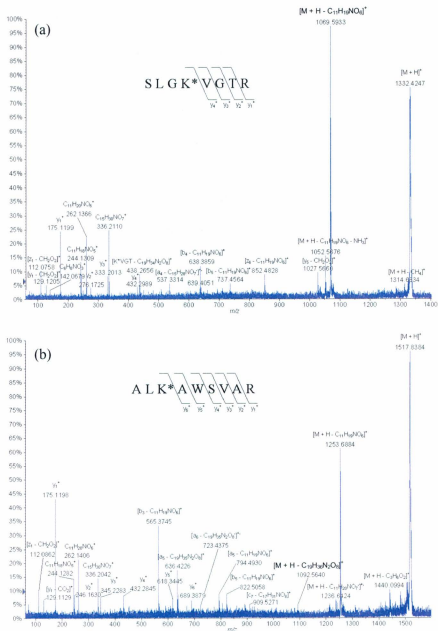


Figure 3.3. MALDI-CID-TOF/TOF-MS/MS spectra of the glycosylated peptide (a) SLGK*VGTR at m/z 1330.72, (b) ALK*AWSVAR at m/z 1514.82 and (c) K*VPQVSTPTLVEVSR at m/z 2153.16.

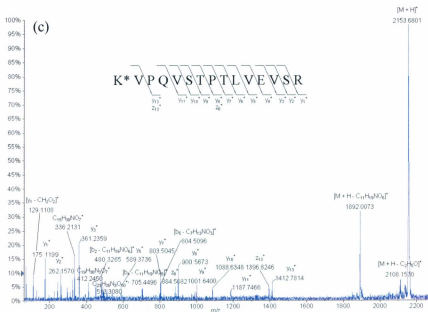


Figure 3.3. (continued)

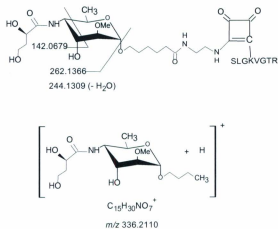


Figure 3.4. Hapten fragments observed during the the MALDI-TOF-MS/MS analysis of the SLGK*VGTR hapten-peptide conjugate at m/z 1330.72.

The MALDI-TOF/TOF-MS/MS analysis of the conjugated peptide with m/z 1514.82 (Figure 3.3(b), Table B.2) allowed us to confirm the glycosylated peptide sequence and the localization of the glycosyl hapten attached to the Lys 235 residue. The direct sequencing of this glycosylated peptide assigned as ALK*AWSVAR (233-241) was afforded by using the diagnostic CID y -product ions. Note that the CID-MS/MS analysis of the ALK*AWSVAR (233-241) gave a series of similar (not identical) product ions derived from glycosylated peptide SLGK*VGTR. These were identified as follows: $[C_{11}H_{18}NO_5]^+$ at m/z 244.13, $[C_{11}H_{20}NO_6]^+$ at m/z 262.14 and $[C_{15}H_{30}NO_7]^+$ at m/z 336.20, as well as the product-ion formed by the loss of glycosyl hapten (as above) fragments: $[M + H - C_3H_6O_2]^+$ at m/z 1440.10, $[M + H - C_{11}H_{19}NO_6]^+$ at m/z 1253.69, $[M + H - C_{11}H_{20}NO_7]^+$ at m/z 1236.64 and $[M + H - C_{19}H_{36}N_2O_8]^+$ at m/z 1092.56. Finally, the presence of product ions of the a -, b - and c - series, produced by the specific cleavage of the conjugated carbohydrate-spacer peptide is, respectively, assigned as: $[b_3 - C_{11}H_{19}NO_6]^+$ at m/z 565.37, $[a_5 - C_{19}H_{35}N_2O_8]^+$ at m/z 636.42, $[a_6 - C_{19}H_{35}N_2O_8]^{++}$ at m/z 723.44, $[a_5 - C_{11}H_{19}NO_6]^+$ at m/z 794.49, $[b_5 - C_{11}H_{19}NO_6]^+$ at m/z 822.51 and $[c_7 - C_{17}H_{31}NO_8]^+$ at m/z 909.53. This result allowed us to confirm the sequence of this glycosylated peptide.

The localisation and the sequence of the glycosylated peptide at m/z 2153.16 were determined by MALDI-TOF/TOF-MS/MS analysis (Figure 3.3(c), Table B.3). It was found that the sequence of this peptide is K*VPQVSTPTLVEVSR (437-451) and that the carbohydrate hapten was conjugated to the Lys 437 residue. As for the previously analyzed glycosylated peptides, we observed product-ions for the glycosyl hapten (as discussed above): $[C_{11}H_{18}NO_5]^+$ at m/z 244.13, $[C_{15}H_{26}N_3O_4]^+$ at m/z 312.21, $[C_{15}H_{30}NO_7]^+$ at m/z 336.21, $[C_{19}H_{30}N_3O_7]^+$ at m/z 412.25, $[C_{23}H_{33}N_3O_9]^{++}$ at m/z 495.26

and $[C_{23}H_{35}N_3O_{10}]^{+*}$ at m/z 513.31. The fragmentation of the glycosyl hapten (as described above) induces also the formation of product ions assigned as follows: $[M + H - C_2H_5O]^+$ at m/z 2108.15 and $[M + H - C_{11}H_{19}NO_6]^+$ at m/z 1892.01. In addition, we also noted the expected series of a- and b-product ions $[b_2 - C_{23}H_{37}N_3O_{10}]^+$ at m/z 226.15, $[a_3 - C_{23}H_{35}N_3O_{10}]^+$ at m/z 297.21, $[b_3 - C_{23}H_{35}N_3O_{10}]^+$ at m/z 325.23, $[b_4 - C_{23}H_{36}N_3O_{10}]^{+*}$ at m/z 452.31, $[b_2 - C_{11}H_{19}NO_6]^+$ at m/z 480.33, $[b_4 - C_{11}H_{19}NO_6]^+$ at m/z 705.45 and $[b_5 - C_7H_{13}NO_3]^+$ at m/z 804.51.

In the MALDI-TOF/TOF mass spectra of the BSA neoglycoconjugate with carbohydrate:protein ratio of 4.3:1, we have identified the major glycosylated peptides at m/z 1330.72, 1514.82 and 2153.16, which had higher intensities than the other non-identified peptides (Figure 3.2). We can, thus, conclude that the synthetic hapten-BSA neoglycoconjugate (4.3:1) carries the carbohydrate hapten preferentially at Lys 455 (RA = 72.6%), Lys 235 (RA = 46.1%) and Lys 437 (RA = 35.9%). It allowed us to conclude that the conjugate consisted of a mixture of isoforms derivatized mainly on the Lys 455, Lys 437 and Lys 235 residues.

For the digested hapten-BSA neoglycoconjugate with a carbohydrate:BSA ratio of 6.6:1, we found that the glycosylated peptides at the Lys 455, Lys 235 and Lys 437 positions had relative abundances of 29.5%, 13.0% and 11.0%, respectively. Similarly, the digest from the neoglycoconjugate with a carbohydrate:protein ratio of 13.2:1 afforded the same series of glycosylated peptides as the latter one, with the respective relative abundances of 12.7%, 10.45% and 4.5%.

The purpose of this work was to identify all the glycosylated peptides present in the digests of the studied neoglycoconjugates (having chemically attached 4 to 13 haptens).

Unfortunately, we were only able to characterize three of these. It is also interesting to note that the carbohydrate conjugation was observed in two different domains of the BSA, that is, in the domain IIA through the binding to the Lys 235 and in the domain IIIA via the conjugation to the Lys 437 and Lys 455.^[22] We presume that these Lys residues belong to those that are considered to be sterically more accessible on the surface of the tridimensional structure. During the MALDI-TOF-MS analysis of the digested BSA neoglycoconjugate, we observed some protonated molecules that did not correspond to any peptides of the digested native BSA. Unfortunately, we were not able to identify them.

Before summarizing conclusions of this investigation, we may pose a simple and obvious question: Since, according to the MALDI-TOF-MS analysis, the BSA-neoglycoconjugate with the highest degree of substitution should carry, on average, ~13 sugars, and each Lys residue carries, presumably, only one sugar-spacer moiety, what happened to the other glycosylated peptides? It is tempting to answer this question with the following rationale. We propose that the carbohydrate-spacer attached to lysine residues alters their chemical properties, and makes the resulting glycosylated peptides less reactive during the enzymatic digestion. This may affect the outcome of the digestion. We, therefore, propose that in order to produce a larger number of diagnostically significant glycosylated peptides, the digestion of BSA-neoglycoconjugate should be carried out with a different peptidase. In addition, these "missing", non-identified protonated molecules may also correspond to glycosylated peptides which are unstable in the gas-phase, and therefore, disintegrate during MALDI-TOF-MS analysis by either in-source or post-source decay. To prevent such phenomena from occurring in our subsequent studies,

electrospray ionization low-energy tandem mass spectrometry is the analytical tool of choice.

Finally, we would like to mention a unique phenomenon that occurred while analyzing this series of neoglycoconjugates by MALDI-TOF/TOF-MS. The single stage MALDI-TOF/TOF-MS of the hapten-BSA conjugate measured with the collision induced dissociation gas in the collision cell turned on resulted in cleavage of some of the covalently attached carbohydrate haptens, with the concomitant release of the BSA ions. This manifested itself by alteration of the average hapten:BSA ratios from 4.3:1, 6.6:1 and 13.2:1, originally found for the glycoconjugates, to 2.0:1, 4.5:1 and 11.5:1, respectively. This appears to indicate that the carbohydrate-carrier bonds are slightly labile to high collision energy, which they are subjected to during the studies by mass spectrometry (Figure 3.5, Table 3.4).

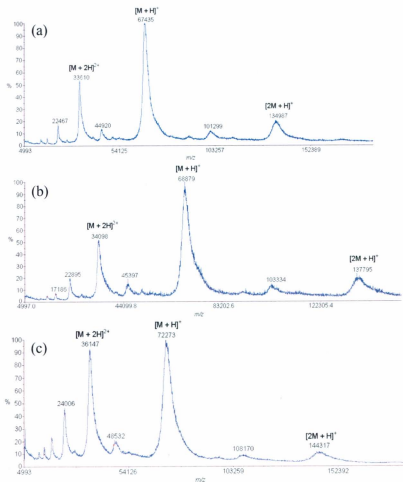


Figure 3.5. MALDI-TOF/TOF-MS analysis of the neoglycoconjugates with hapten:BSA ratios 4:1 (a), 7:1 (b) and 13:1 (c), with the collision cell switched on.

Table 3.4. Molecular masses of BSA and hapten-BSA conjugates and their calculated hapten:BSA ratios, obtained during the MALDI-TOF-MS analysis of the glycoconjugate with the collision cell switched on.

	[M + 2H] ²⁺	[M + H] ⁺	[2M + H] ⁺	Recalculated Hapten:BSA ratio
BSA standard	33,115	66,310	133,019	-
Hapten:BSA 4.3:1	33,610	67,435	134,987	2.0:1
Hapten:BSA 6.6:1	34,098	68,879	137,795	4.5:1
Hapten:BSA 13.2:1	36,147	72,273	144,317	11.5:1

3.4. Conclusion

The detailed structural analysis of neoglycoconjugates is an important step in the verification of the structure of conjugate vaccines before their use as pharmaceuticals. In this study, a carbohydrate hapten-BSA neoglycoconjugate was analyzed by MALDI-TOF/TOF-MS in order to confirm its structure and to localize the sites of the chemical attachment of the carbohydrate hapten onto the carrier protein. To explore the feasibility of the aforementioned task, three different carbohydrate-BSA neoglycoconjugates were subjected to MALDI-TOF/TOF-MS analysis of their proteolytic digests. The conjugates, which were obtained with different carbohydrate:protein ratios (4.3:1, 6.6:1 and 13.2:1), were prepared by chemical attachment of the carbohydrate hapten to the protein carrier. We know that the product-ion scans, measured with the MALDI-TOF/TOF-MS/MS, occur with high energy collision induced dissociation. This high-energy precursor fragmentation promotes the primary fragmentations of the carbohydrate hapten, resulting in CID spectra containing amino acid sequences attached to fragments of the carbohydrate. As a result, we were able to identify the conjugation sites involving three different lysine residues (Lys 235, Lys 437 and Lys 455). Although only few points of glycation could be determined, the identification of the y-series ions was very useful for

the peptide sequencing. The series of a- and b-ions confirmed the sequence of the conjugated peptides, although they included product ions containing fragments of the hapten. Identification of the other unknown peptides will be carried out in the future, using a MALDI-QqTOF-MS/MS hybrid instrument using lower-energy collision dissociation and FTICResonance mass spectrometer with an electron capture dissociation cell, which is known to produce mainly c- and z-ions during the MS/MS analysis of peptides.

Appendix B. Supplementary data

Supplementary data are associated with this chapter (page 243)

References

- [1] C. Galanos, O. Luderitz. in *Handbook of Endotoxins, 1, Chemistry of Endotoxins*, Ed by Rietschel ET. Elsevier, Amsterdam, The Netherlands, **1984**, p. 46.
- [2] E. T. Rietschel, H. W. Wollenweber, H. Brade, U. Zahring, B. Linder, U. Seydel, H. Bradaezek ,G. Barnickel, H. Labischinski, P. Giesbrecht. in *Handbook of Endotoxins, 1, Chemistry of Endotoxins*, Ed by Rietschel ET. Elsevier, Amsterdam, The Netherlands, **1984**, p. 187.
- [3] O. Holst, H. Brade. in *Bacterial Endotoxic Lipopolysaccharides, 1, Molecular Biochemistry and Cellular Biology*, Ed by Morrison DC, Ryan JL. CRC Press, Boca Raton, Florida, USA, **1992**, p. 135.
- [4] J. B. Robbins, R. Schneerson, S. C. Szu, V. Pozsgay. Bacterial polysaccharide-protein conjugate vaccines. *Pure Appl. Chem.* **1999**, 71, 745.
- [5] C. R. H. Raetz. Biochemistry of Endotoxins. *Ann. Rev. Biochem.* **1990**, 59, 129.
- [6] J. Kubler-Kielb, E. Vinogradov, G. Ben-Menachem, V. Pozsgay, J. B. Robbins, R. Schneerson. Saccharide/protein conjugate vaccines for *Bordetella* species: preparation of saccharide, development of new conjugation procedures, and physico-chemical and immunological characterization of the conjugates. *Vaccin.* **2008**, 26, 3587.
- [7] V. Pozsgay. Recent developments in synthetic oligosaccharide-based bacterial vaccines. *Curr. Top. Med. Chem.* **2008**, 8, 126.
- [8] C. Grandjean, A. Boutonnier, B. Dassy, J. M. Fournier, L. A. Mulard. Investigation towards bivalent chemically defined glycoconjugate immunogens prepared from aciddetoxified lipopolysaccharide of *Vibrio Cholerae* O1, serotype Inaba. *Glycoconj. J.* **2009**, 26, 41.

- [9] J. H. Banoub, D. H. Shaw, N. A. Nakhla, H. J. Hodder. Synthesis of glycoconjugates derived from various lipopolysaccharides of the *Vibrionaceae* family. *Eur. J. Biochem.* **1989**, *179*, 651.
- [10] A. Ariosa-Alvarez, A. Arencibia-Mohar, O. Madrazo-Alonso, L. Garcia-Imia, G. Siera-Gonzalez, V. Verez-Bencomo. Synthesis of the *Vibrio cholerae* O1 Ogawa and Inaba terminal disaccharides with dioxolane-type spacers and their coupling to proteins. *J. Carbohydr. Chem.* **1998**, *17*, 1307.
- [11] T. K. Wade, R. Saksena, J. Shiloach, P. Kováč, W. F. Wade. Immunogenicity of synthetic saccharide fragments of *Vibrio cholerae* O1 (Ogawa and Inaba) bound to Exotoxin A. *FEMS Immunol. Med. Microbiol.* **2006**, *48*, 237.
- [12] A. Chernyak, S. Kondo, T. K. Wade, M. D. Meeks, P. M. Alzari, J. M. Fournier, R. K. Taylor, P. Kovac, W. F. Wade. Induction of protective immunity by synthetic *Vibrio cholerae* hexasaccharide derived from *V. cholerae* O1 Ogawa lipopolysaccharide bound to a protein carrier. *J. Infect. Dis.* **2002**, *185*, 950.
- [13] S. Chatterjee, K. Chaudhuri. Lipopolysaccharides of *Vibrio cholerae*. I. Physical and chemical characterization. *Biochem. Biophys. Acta.* **2003**, *1639*, 65.
- [14] E. T. Rietschel, L. Brade, B. Lindner, U. Zahringer. in *Bacterial Endotoxic Lipopolysaccharides*, *1*, Ed by Morrison DC, Ryan JL. CRC Press. Boca Raton, FL, **1992**, p 3.
- [15] P. A. Manning, U. H. Stroecher, R. Morona. in *Vibrio cholerae and Cholera: Molecular to Global Perspectives*, Ed by Wachsmuth IK, Blake PA, Olsvik O. *American Society for Microbiology*, Washington, DC, **1994**, p. 77.

- [16] W. E. Dick Jr., M. Beurret. in *Conjugate Vaccines, 10*, J. M. Cruse, R. E. Lewis Jr. (eds), Krager, Basel, **1989**, p. 48.
- [17] A. D. McNaught. International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology. Joint commission on biochemical nomenclature. Nomenclature of carbohydrates. *Carbohydr. Res.* **1997**, 297, 1.
- [18] L. Kenne, B. Lindberg, P. Unger, B. Gustafsson, T. Holme. Structural studies of the *Vibrio cholerae* O-antigen. *Carbohydr. Res.* **1982**, 100, 341.
- [19] K. Hisatsune, S. Kondo, Y. Isshiki, T. Igushi, Y. Haishima. Occurrence of 2-OMethyl-N-(3-Deoxy-L-glycero-tetronyl)-D-perosamine (4-amino-4,6-dideoxy-D-mannopyranose) in lipopolysaccharide from Ogawa but not from Inaba O forms of O1 *Vibrio cholerae*. *Biochem. Biophys. Res. Commun.* **1993**, 190, 302.
- [20] Y. Isshiki, S. Kondo, Y. Haishima, T. Iguchi, K. Hisatsune. Identification of N-3-hydroxypropionyl-2-O-methyl-D-perosamine as a specific constituent of the lipopolysaccharide from *Vibrio* bio-serogroup 1875 which has Ogawa antigen factor B of *Vibrio cholerae* O1. *J. Endotoxin Res.* **1996**, 3, 143.
- [21] C. Hilger, F. Grigioni, C. De Beaufort, G. Michel, J. Freilinger, F. Hentges. Differential binding of IgG and IgA antibodies to antigenic determinants of bovine serum albumin. *Clin. Exp. Immunol.* **2001**, 123, 387.
- [22] B. X. Huang, H. Y. Kim, C. Dass. Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, 15, 1237.

- [23] C. E. Costello, J. M. Contado-Miller, J. F. Cipollo. A glycomics platform for the analysis of permethylated oligosaccharide alditols. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1799.
- [24] B. Domon, C. E. Costello. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **1988**, *27*, 1534.
- [25] B. Domon, C. E. Costello. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, *5*, 397.
- [26] S. Carganico, P. Rovero, J. A. Halperin, A. M. Panini, M. Chorev. Building blocks for the synthesis of post-translationally modified glycated peptides and proteins. *J. Pept. Sci.* **2009**, *15*, 67.
- [27] J. M. Ames. Mass spectrometry to detect the site specificity of advanced glycation/lipoxidation end-product formation on protein: some challenges and solutions. *Biochem. Soc. Trans.* **2008**, *36*, 1051.
- [28] X. Zhang, K. F. Medzihradsky, J. Cunningham, P. D. K. Lee, C. L. Rognnerud, C. N. Ou, P. Harmatz, H. E. Witkowska. Characterization of glycated hemoglobin in diabetic patients: usefulness of electrospray mass spectrometry in monitoring the extent and distribution of glycation. *J. Chromatogr. B* **2001**, *759*, 1.
- [29] S. Guedes, R. Vitorino, M. R. M. Domingues, F. Amado, P. Domingues. Mass spectrometry characterization of the glycation sites of bovine insulin by tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1319.

- [30] V. P. Kamath, P. Diedrich, O. Hindsgaul. Use of diethyl squarate for the coupling of oligosaccharide amines to carrier proteins and characterization of the resulting neoglycoproteins by MALDI-TOF mass spectrometry. *Glycoconj. J.* **1996**, *13*, 315.
- [31] A. J. Dambra, J. E. Baugher, P. E. Concannon, R. A. Pon, F. Michon. Direct and indirect methods for molar-mass analysis of fragments of the capsular polysaccharide of *Haemophilus influenzae* type b. *Anal. Biochem.* **1997**, *250*, 228.
- [32] P. Kovac. in *Synthetic Oligosaccharides. Indispensable Probes in the Life Sciences*, Ed.; ACS Symposium Series 560; American Chemical Society: Washington, DC, 1994.
- [33] V. Pozsgay. Synthesis of glycoconjugate vaccines against *Shigella dysenteriae* type 1. *J. Org. Chem.* **1998**, *63*, 5983.
- [34] R. Saksena, A. Chernyak, A. Karavanov, P. Kovác. Conjugating low molecular mass carbohydrates to proteins. I. Monitoring the progress of conjugation. *Meth. Enzymol.* **2003**, *362*, 125.

CHAPTER 4: Determination of the glycation sites of *Bacillus anthracis* neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-tandem mass spectrometry

This chapter has been published: Farid Jahouh,¹ Shu-jie Hou,² Pavol Kováč² and Joseph H. Banoub^{1,3*} *J. Mass Spectrom.* **2011**, *46*, 993.

¹Department of Chemistry, Memorial University of Newfoundland, Saint John's, NL,

A1B 3V6, Canada

²NIDDK, LBC, National Institutes of Health, Bethesda, MD 20892-0815, USA

³Department of Fisheries and Oceans Canada, Science Branch, Special Projects, Saint

John's, NL, Canada, A1C 5X1

Abstract

We present herein an efficient mass spectrometric method for the localization of the glycation sites of a model neoglycoconjugate vaccine model formed by a construct of the tetrasaccharide side chain of the *Bacillus anthracis* exosporium and the protein carrier BSA. The glycoconjugate was digested with both trypsin and GluC V8 endoproteinase and the digests were then analyzed by MALDI-TOF/TOF-CID-MS/MS and nano-LC-ESI-QqTOF-CID-MS/MS. The sequences of the unknown peptides analyzed by MALDI-TOF/TOF-CID-MS/MS, following digestion with the GluC V8 endoproteinase, allowed us to recognize five glycopeptides whose glycation occupancies were respectively on Lys 235, Lys 420 and Lys 498. Similarly, the same analysis was performed on the tryptic digests which permitted us to recognize the glycation sites on Lys 100 and Lys 374. In addition, we have also used LC-ESI-QqTOF-CID-MS/MS analysis for the identification of the tryptic digests. However, this analysis identified a higher number of glycopeptides than would be expected from a glycoconjugate composed from carbohydrate:protein ratio of 5.4:1, which would have resulted in a glycation occupancy of 18 specific sites. This discrepancy was due to the large number of glycoforms formed during the synthetic carbohydrate-spacer-carrier protein conjugation. Likewise, the LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digest also identified 17 different glycation sites on the glycoconjugate.

4.1. Introduction

Bacterial LPS have been extensively studied during the last decades and their use in the synthesis of neoglycoconjugates as potential vaccines has been successful.^[1-7] Different parameters have been found to influence the efficacy of a glycoconjugate vaccine, such as the saccharide size, the average number of saccharide chains per conjugate molecule, nature of the carrier and the distance between the saccharide and the protein in the formed glycoconjugate.^[8-11] Until very recently, the covalent attachment sites of the antigenic saccharide to the protein in all neoglycoconjugates have not been determined. We were the first to attempt such an investigation and have unraveled some of the glycation sites in the neoglycoconjugates made from BSA and the *O*-PS of *Vibrio cholerae* O1, serotype Ogawa.^[12]

Bacillus anthracis belongs to the Gram-positive bacterium family and is the cause of anthrax, a disease which can affect animals and humans.^[13] The bacterium is known to form endospores^[14] which, when mature, are resistant to extreme conditions (temperature, radiations, desiccation, physical damage, harsh chemicals).^[15] The currently increased interest in anthrax is because some forms of *Bacillus anthracis*, which are the etiologic agent of anthrax,^[16] including anthrax spores, can be used as biological weapons. Several studies have been carried out on the *B. anthracis* to understand its pathogenesis, to discover new markers for its detection, and also to identify constituents of its spore that could be targeted for its inactivation.^[17]

Different synthetic vaccines have been developed against anthrax; these were based on the immunogenic conjugates of the capsular polypeptide (polyglutamic acid) of *B. anthracis*, and an anthrax capsular vaccine has also been developed.^[18]

A better understanding of the anthrax spores composition and structure was necessary before a vaccine for anthrax could be developed based on a component of anthrax spores. Recently, the structure of the tetrasaccharide side chain of the collagen like region of the major glycoprotein of the *B. anthracis* exosporium was determined by Daubenspeck *et al.*^[19] The upstream terminal of the tetrasaccharide was found to be the new sugar anthrose [4,6-dideoxy-4-(3-hydroxy-3-methylbutyramido)-2-O-methyl-D-glucopyranose]. By comparing the NMR data of the synthetic tetrasaccharide^[20] with those recorded^[19] for the tetrasaccharide isolated from *B. anthracis* exosporium, it was confirmed that the structure of the tetrasaccharide was correct as proposed.^[19] Kováč *et al.* synthesized the tetrasaccharide side chain and linked it to the BSA using squaric acid chemistry.^[21] They obtained synthetic glycoconjugates having an average carbohydrate:protein ratio of 3.5:1 and 5.7:1. Figure 4.1 displays a schematic representation of the general structure of the carbohydrate-protein neoglycoconjugate construct from the tetrasaccharide side chain of the *B. anthracis* exosporium.

The analysis and the localization of the glycation sites in carbohydrate-protein neoglycoconjugates is a very important feature for the quality control of the conjugate vaccines. Different strategies have been developed for the analysis of glycoproteins, mainly based on the MS analysis of enzymatic digests.^[22,23] In a previous study on the determination of the glycation sites in synthetic neoglycoconjugate models of *Vibrio cholerae* O1, we have shown that out of the 59 possible glycation sites on BSA, we were only able to characterize three, when using a standard method which consists on the trypsin digestion of the glycoconjugate. We hypothesized that this protease may not be the best to release all the glycated peptides from this neoglycoconjugate.^[12] In addition,

the unraveling of some of the glycation sites in this neoglycoconjugate was achieved by performing MS/MS analysis using a MALDI-TOF/TOF-MS/MS instrument, with the benefit of high-energy collision dissociation.^[12]

In this manuscript, we present an efficient method for the localization of the glycation sites of the neoglycoconjugate formed by the antigenic tetrasaccharide side chain of the *B. anthracis* exosporium covalently attached to the BSA protein carrier. The digestion of the glycoconjugate was carried out with two different proteases: trypsin and GluC V8. The analysis and sequencing of the resulting peptides and glycopeptides was performed separately by using two different state-of-the art mass spectrometric and tandem mass spectrometric methods.

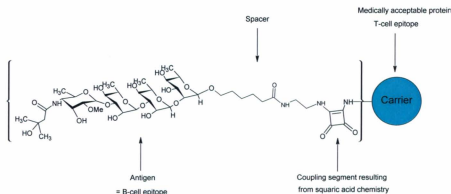


Figure 4.1. Schematic representation of the general structure of carbohydrate-protein constructs from oligosaccharide fragments of the antigenic tetrasaccharide side chain of the *B. anthracis* exosporium.

4.2. Materials and Methods

4.2.1. Preparation of the hapten-BSA neoglycoconjugate

The synthesis of the hapten-BSA glycoconjugate has been described previously.^[21] Thus, the synthetic 2-*O*-methyl- β -D-glucopyranosyl-(1 \rightarrow 3)- α -L-rhamnopyranosyl-(1 \rightarrow 3)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranoside corresponding to the *B. anthracis* exosporium tetrasaccharide was first treated with ethylenediamine to give the 2-aminoethyl amide, which, when treated with diethyl squarate afforded the squaric acid monoethylester. The conjugation of this hapten to BSA (Sigma Aldrich, Saint Louis, MO, USA) in 0.5 M borate buffer (pH 9.0) was followed by SELDI-TOF-MS,^[27] which made it possible to obtain the hapten-BSA glycoconjugates, with a predetermined carbohydrate:protein ratio of \sim 5:1. Isolation of the neoglycoconjugate was carried out by ultrafiltration with the aid of a centrifugal device with a cut-off 30,000 Da.

4.2.2. Digestion

The digestions of the hapten-BSA glycoconjugates were carried out with trypsin and Glu-C V8 protease (Sigma Aldrich, Saint Louis, MO, USA). Thus, 100 μ g of the glycoconjugate (1.4 nmol) was dissolved in a mixture of 0.1% RapiGest SF Surfactant (1 μ g, Waters, USA) in 50 mM of NH_4HCO_3 (100 μ L) at a pH of 8.0 and reduced by treatment with 2 μ L of 10 mM dithiothreitol (Sigma Aldrich, Saint Louis, MO, USA) for 30 min at room temperature, followed by alkylation with 2 μ L of a 50 mM iodoacetamide (Sigma Aldrich, Saint Louis, MO, USA) for 1 h at room temperature. A portion (50 μ g, 0.7 nmol) of the BSA glycoconjugate was digested with trypsin using a 20 ng/mL of trypsin dissolved in NH_4HCO_3 (50 mM, 1 mL) at a trypsin:glycoprotein ratio of 1:25

(w/w) and incubated at 37 °C overnight with shaking. The other 50 µg (0.7 nmol) of the BSA glycoconjugate was digested using the GluC V8 endoprotease:glycoprotein ratio of 1:25 (w/w) and incubated at 37 °C overnight with shaking. The sample was then dried under vacuum and the residue was dissolved in 20 µL of 1% acetic acid (Sigma-Aldrich, Oakville, ON, Canada). An aliquot of each sample (10 µL) was then cleaned up using ZipTip C18 (Millipore, Bedford, MA, USA) before mass spectral analysis.

4.2.3. MALDI-TOF/TOF-MS and MALDI-TOF/TOF-CID-MS/MS analyses

Mass spectral analyses were carried out with an ABI 4800 equipped with TOF-TOF optics (Applied Biosystems Foster City, CA) and a 200-Hz frequency Nd:YAG laser. α -Cyano-4-hydroxycinnamic acid was used as matrix for the analysis of BSA, hapten-BSA conjugate, and the peptides resulting from their digestions, with an average of 5000 to 8000 laser shots per spectra. Basically, 1 µL of a 20 mg/mL solution of α -CHCA (dissolved in acetone, 0.1% TFA) was spotted on the MALDI plate and dried at room temperature. Then, an aliquot of 1 µL of sample was spotted on the top of the dried matrix and let to dry before the MALDI-MS experiments.

The tandem mass spectrometric analyses were achieved with air as collision gas and at collision energy of 1 kV, which correspond to the difference between the accelerating potential (8 kV) and the floating cell (7 kV).^[28] The internal calibration of the digested glycoproteins and BSA was performed in the range m/z 600-3200 using known peptides of the BSA.

4.2.4. LC-ESI-QqTOF-CID-MS/MS analysis

The peptides were separated on a DIONEX UltiMate3000 Nano LC System (Germering, Germany). 250 fmol of digested glycoprotein was dissolved in 0.1% TFA and loaded onto a precolumn (300 μm i.d. \times 5 mm, C18 PepMap100, 5 μm (LC Packing, Sunnyvale, CA)) in order to desalt and concentrate the sample. After their elution from the precolumn, the mixture of peptides and glycopeptides was separated on a nanoflow analytical column (75 μm i.d. \times 15 cm, C18 PepMap 100, 3 μm , 100 Å, (LC Packing, Sunnyvale, CA)) at a flow rate of 180 nL/min. The mobile phase eluents used were composed of 0.1% FA/0.01% TFA/2% ACN (A) and 0.08% FA/0.008% TFA/98% ACN (B). The elution started with 0% B for 10 min, followed by a gradient of 0-60% B in 55 min and 60-90% B in 3 min and was kept at 90% B for 3 min. The tandem mass spectrometry analysis of the eluted peptides and glycopeptides was accomplished using an Applied Biosystems API-QSTAR XL quadrupole orthogonal time-of-flight (QqTOF)-MS/MS hybrid tandem mass spectrometer (Applied Biosystems International-MDS Sciex, Foster City, CA, USA) equipped with a nano-electrospray source (Protana XYZ manipulator) which produces the electrospray through a PicoTip needle (10 μm i.d., New Objectives, Woburn, MA, USA) carrying a voltage of 2400 V. The TOF analyzer was calibrated using a renin solution (1 pmol/ μL) and looking for the ions at m/z 586.9815 and m/z 879.9723. The collision energies used during the CID-MS/MS analyses were determined automatically using the Information Dependent Acquisition method integrated in the Analyst software.

4.3. Results

As mentioned earlier, the aim of this study was to localize the glycation sites for the neoglycoconjugate composed of the antigenic tetrasaccharide side chain of *B. anthracis* exosporium covalently attached to the BSA protein carrier. In this rationale, we have aimed to search the best method for the characterization of the glycopeptides. The digestion of the glycoconjugate was carried out with two different proteases: trypsin and GluC V8. Although trypsin cleaves proteins at the carboxyl side of lysine and/or arginine residues,^[24] GluC V8 endoproteinase is known to cleave protein at the C-terminus of the peptide bond on glutamic acid residues at pH 4.0 and on both glutamic acid and aspartic acid residues at pH 7.8.^[25,26]

Furthermore, for the analysis and sequencing of the resulting peptides and glycopeptides, we have used two different mass spectrometric approaches. The first one involved the analysis of the digests by MALDI-TOF/TOF-MS (MALDI-MS), followed by peptides sequencing using high collision energy MALDI-TOF/TOF-CID-MS/MS (MALDI-MS/MS). The second approach involved analysis of the mixed peptide digests by LC-ESI-QqTOF-MS (LC-MS) followed by peptide sequencing using low-energy collision induced dissociation CID-QqTOF-MS/MS analysis (LC-MS/MS). A scheme, indicating the general strategy used in this manuscript, is illustrated in Figure 4.2.

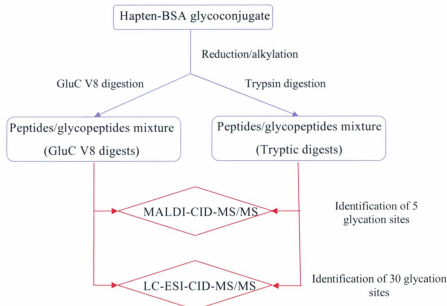


Figure 4.2. General strategy applied for the mass spectrometry determination of the glycation sites on the hapten-BSA glycoconjugate.

4.3.1. MALDI-TOF/TOF-MS analysis of BSA and the hapten-BSA neoglycoconjugates

Single stage MALDI-MS analysis of the BSA and the carbohydrate hapten-BSA glycoconjugate allowed us to determine the average number of carbohydrate-spacer covalently attached to the BSA by comparison of the observed molecular weight of the glycoconjugate to the molecular weight of the native BSA. Thus, the MALDI-MS analysis of the hapten-BSA glycoconjugate allowed us to identify the protonated molecule $[M + H]^+$ at m/z 71,448.36 and $[M + 2H]^{2+}$ at m/z 35,730.08 (Figure C.1, Appendix C). Using these data, we calculated the molecular mass of the hapten-BSA glycoconjugate which was 71,448.36 Da. Finally, the comparison of the observed molecular weight of the BSA and the hapten-BSA enabled us to calculate the average

hapten:BSA ratio on the glycoconjugate. As previously reported, during the conjugation, the squaric acid is covalently linked to the ϵ -amino groups of the lysine residues in BSA with loss of a molecule of ethanol.^[9,12,21] Knowing the molecular mass of the tetrasaccharide-squaric acid unit (950.43 Da), we were able to calculate the carbohydrate hapten:BSA ratio which was found to be 5.4:1. The glycated peptides on the BSA were identified by the peptide mass fingerprinting (PMF) experiment on the hapten-BSA conjugate digests and the CID-MS/MS analysis of the glycated peptides permitted to localize the glycation site.

4.3.2. PMF analysis using MALDI-TOF/TOF-MS of the digested hapten-BSA neoglycoconjugates

MALDI-MS analysis was carried out on the digested (trypsin and GluC V8) hapten-BSA glycoconjugate and the obtained MS spectra were submitted to the Mascot library to identify by PMF the peptides matching to the BSA (Figure 4.3, Tables C.1 and C.2). The following parameters were applied for the Mascot library searching: carbamidomethyl (C) as fixed modifications, oxidation (M) as variable modifications, a peptide mass tolerance of ± 0.2 Da and a maximum of missed cleavages of 1.

The Mascot report of the tryptic and GluC V8 digests spectra gave a match to two isoforms of the Serum Albumin protein from the *Bos taurus* species, namely, the serum albumin precursor (gi|1351907) and the serum albumin (gi|74267962). Thus, we postulated that the obtained peptides that were not a match with the peptides of the BSA had to be conjugated with the hapten carbohydrate. Consequently, these were analyzed by MALDI-MS/MS analysis.

The MALDI-MS analysis of the carbohydrate hapten-BSA glycoconjugate digested with trypsin (Figure 4.3(a) Table C.1) afforded only three identifiable glycopeptides having different glycation sites. These diagnostic three glycopeptides corresponded to the BSA peptides whose molecular masses have increased by 950.43 Da. These were verified identified to have the following structures: ALK*AWSVAR at m/z 1951.0130, VTK*CCTESLVNR at m/z 2416.1372, and QNCDQFEK*LGEYGFQNALIVR at m/z 3478.6429 (glycation represented with an asterisk on the lysine residue).

The MALDI-MS analysis of the glycopeptides resulting from the digestion of the hapten-BSA glycoconjugate with the GluC V8 protease (Figure 4.3(b), Table C.2) afforded only three glycopeptides: LCK*VASLRE at m/z 2025.0186, YAVSVLLRLAK*E at m/z 2311.2476 and YAVSVLLRLAK*EYE at m/z 2603.3484.

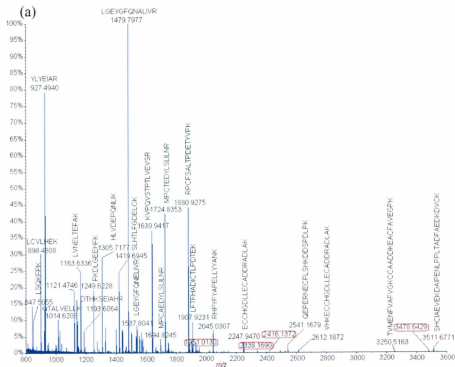


Figure 4.3. MALDI-MS analysis of the glycoconjugate trypsin digests (a) and GluC V8 endoproteinase digests (b).

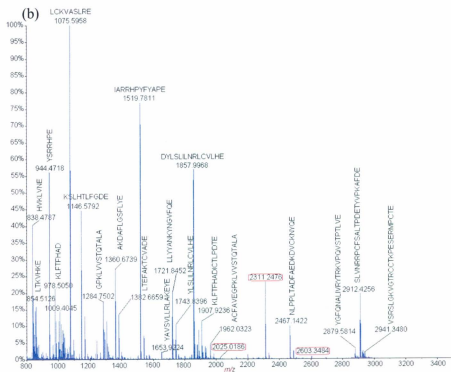


Figure 4.3. (continued)

4.3.3. High-energy MALDI-TOF/TOF-CID-MS/MS analyses of the glycosylated peptides

4.3.3.1. MALDI-TOF/TOF-CID-MS/MS analyses of the glycosylated peptides obtained after trypsin digestion of the carbohydrate hapten-BSA glycoconjugate

The sequence and the determination of the glycosylation sites of the trypsin obtained glycopeptides were verified by MS/MS. Thus, we carefully analyzed by MALDI-MS/MS analyses the precursor protonated molecules $[M + H]^+$ extracted from the following glycopeptides: ALK*AWSVAR (Lys 235) at m/z 1951.0130 presented in Figure 4.4(a)

(Table C.3); VTK*CTESLVNR at m/z 2416.1372 (Figure C.2), and QNCDQFEK*LGEYGFQNALIVR at m/z 3478.6429 (Figure C.3).

During the high-energy CID-MS/MS analysis of the glycopeptide ALK*AWSVAR (Lys 235) at m/z 1951.0130 (Figure 4.4(a), Table C.3), we observed a series of consecutive losses due to the various carbohydrate fragments. These losses seem to indicate that during CID-MS/MS analyses, the glycosyl hapten fragmented in the gas-phase, and this permits the unambiguous identification and localization of the glycation site.

For the nomenclature of the glycopeptide product ions obtained, we have used the nomenclature established by Domon and Costello for the carbohydrate portion as: A, B, C, X, Y and Z.^[29] Thus, we observed that the product ions generated by the fragmentation of the carbohydrate hapten, occurred by the successive losses of either each of the monosaccharide B₁ (-259 Da), disaccharide B₂ (-405 Da), trisaccharide B₃ (-551 Da) and/or the tetrasaccharide B₄ (-697 Da), leading respectively to the formation of the following product ions: Y₄⁺ at m/z 1691.8829, Y₃⁺ at m/z 1545.8376, Y₂⁺ at m/z 1399.7768 and Y₁⁺ at m/z 1253.7010. We also noted that the difference between Y₃⁺ and Y₂⁺, Y₂⁺ and Y₁⁺, Y₁⁺ and Y₀⁺ was 146 Da (Figure 4.4(a), Table C.3), and this difference corresponded to one α -L-rhamnopyranosyl unit. We therefore propose in this study, that the loss of a glycosyl unit represents a diagnostic fragmentation signature of the synthetic tetrasaccharide. This attribute is extremely important for the confirmation of the identity of the carbohydrate hapten covalently linked to the peptide. In addition, different carbohydrate product ions resulting from other fragmentation routes were also identified and these were assigned as follows: C₁₇H₂₆N₃O₄⁺ at m/z 336.1242, B₁⁺ at m/z 260.0739,

$[B_1 - H_2O]^+$ at m/z 242.0709, $^{2,5}A_1^+$ at m/z 230.0719, $C_{11}H_{16}N_3O_2^+$ at m/z 222.0670 and $[C_1 - C_5H_{11}NO_2]^+$ at m/z 159.0486 [Figure 4.4(b)].

To minimize the ambiguity in the nomenclature used in this study, we have also used the pre-existing peptides nomenclature involving the detection of the x-, y-, z-ions and the a-, b-, c-ions for the individual peptide sequence.^[30,31] In consequence, it is important to point out that the b- and y-product ions obtained, included the glycosylated lysine, which have lost part of the carbohydrate moiety. However, it is important to comprehend that these peptides still preserved the portion of the attached spacer-squaric acid chain. The following are examples of the b- and y-product ions that we have observed: $[b_3 - B_4]^+$ at m/z 565.2661, $[b_4 - B_4]^+$ at m/z 636.2715 and $[y_7 - B_2]^+$ at m/z 1361.7265.

High-energy CID-MS/MS analysis of the $[M + H]^+$ precursor ion at m/z 2416.1372 of the glycopeptide VTK*CTESLVNR (Figure C.2) permitted us to pinpoint the tetrasaccharide hapten location site on the Lys 498. This CID-MS/MS analysis also showed the consecutive losses of the glycosyl portions of the tetrasaccharide B_{1-4} to afford the product ions Y_{0-3} and other oligosaccharide product ions (Figure C.2). Finally, we were able to identify the following glycopeptide as QNCDQFEK*LGEYGFQNALIVR at m/z 3478.6429 (Figure C.3) which contained the carbohydrate hapten covalently attached to the Lys 420 residue.

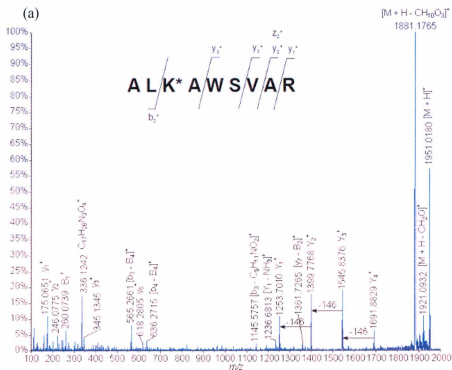


Figure 4.4. (a) MALDI-TOF/TOF-MS/MS spectra of the glycosylated peptide ALK*AWSVAR (Lys 235) at m/z 1951.0130. (b) Different product ions involving the fragmentation of the carbohydrate hapten observed during the MALDI-TOF/TOF-MS/MS analysis of the glycosylated peptide ALK*AWSVAR (Lys 235) at m/z 1951.0130.

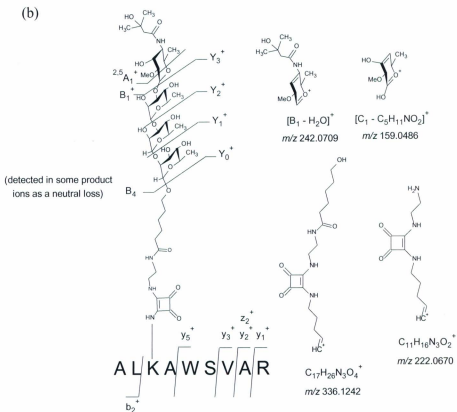


Figure 4.4. (continued)

It is noteworthy to mention that during high-energy CID-MS/MS analysis, we were subjected to a small amount of flawed results. For instance, what we presumed to be a glycosylated peptide at m/z 2338.1690, turned unfortunately to be the intact entire non-glycosylated peptide [MPCTEDYLSLILNR]. This discrepancy was most probably due to the fact that during MALDI-MS/MS analysis the instrument ion gate selection of the precursor ion allowed also the transmission of non-glycosylated peptide which was isobaric to the selected precursor ion at m/z 2338.1690. In addition, it has been recently reported that the ions produced in the MALDI source cause post source decay^[32,33] and that they may overlap the product ions resulting from the high energy CID analysis of selected precursor ions, which complicate the product ions spectra.^[34]

4.3.3.1. MALDI-TOF/TOF-CID-MS/MS analyses of the glycosylated peptides obtained after the GluC V8 digestion of the carbohydrate hapten-BSA conjugate

The digested GluC V8 endoproteinase glycosylated peptides were detected by single stage MALDI-MS analysis and analyzed by MALDI-MS/MS. The high energy CID-MS/MS analysis of the precursor ion $[M + H]^+$ at m/z 2025.0186 extracted from the glycopeptide LCK*VASLRE (Figure C.4, Table C.4) allowed us to confirm the sequence of the unglycosylated complete glycopeptide ion and to localize the carbohydrate occupancy on the covalently attached Lys 100 residue. Once more, the protonated glycopeptide precursor ion at m/z 2025.0186 afforded various peptide product ions containing the full intact peptide portion, which have lost sections of the intact tetrasaccharide hapten. These product ions were identified as: Y_3^+ at m/z 1765.8924, Y_2^+ at m/z 1619.8220, Y_1^+ at m/z 1473.7520, Y_0^+ at m/z 1327.7091 and $[Y_0 - NH_3]^+$ at m/z 1310.6732. The sequence coverage of this glycopeptide is shown in Figure C.4 and, as expected, some peptide

product ions were observed with the loss of the entire tetrasaccharide B₄ moiety: [b₆ - B₄]⁺ at *m/z* 911.4274, [b₅ - B₄]⁺ at *m/z* 824.3805, [b₄ - B₄]⁺ at *m/z* 753.3832, [b₃ - B₄]⁺ at *m/z* 654.3269 and [K⁺V - B₄]⁺ at *m/z* 452.3055. In addition, the following diagnostic carbohydrate-spacer product ions created by fragmentation of the tetrasaccharide spacer portion were also detected: C₁₇H₂₆N₃O₄⁺ at *m/z* 336.2023, B₁⁺ at *m/z* 260.1505, [B₁ - H₂O]⁺ at *m/z* 242.1599, C₁₁H₁₆N₃O₂⁺ at *m/z* 222.1627, [C₁ - C₃H₇O]⁺ at *m/z* 202.1093 and C₇H₁₃O₂⁺ at *m/z* 129.1094.

High-energy CID-MS/MS analysis of the precursor ions [M + H]⁺ at *m/z* 2311.2476 extracted from the glycopeptide YAVSVLLRLAK*E (Figure C.5) and at *m/z* 2603.3484 for the glycopeptide YAVSVLLRLAK*EYE (Figure C.6) revealed that the second glycation site was located on the Lys 374 residue. To sum up, the MALDI-MS/MS analysis of the digested carbohydrate-BSA conjugate with trypsin and GluC V8 allowed us to identify five glycation sites: Lys 100, Lys 235, Lys 374, Lys 420 and Lys 498. It is crucial to point out that the detection of carbohydrate product ions are diagnostic of the presence of the glycated peptide and also allow the validation of the location of the covalently attached carbohydrate hapten sites. However, we report herein a conjecture on the uses of the single stage MALDI-MS analyses, which in our opinion can be attributed to the low sensitivity of detection of the mixture of glycopeptides and accordingly, some glycated peptides were not detected at all. Moreover, the low precursor ion selectivity of the linear TOF in addition to the overlapping of the PSD and CID fragment ions can complicate the interpretation of the product ions spectrum.^[34,35]

4.3.4. LC-ESI-QqTOF-CID-MS/MS analysis of the digested carbohydrate-BSA conjugate

In this rationale, as mentioned before, we have used a second approach, involving the analysis of the peptide digests by LC-ESI-QqTOF-MS (LC-MS) followed by peptide sequencing using low-energy CID-QqTOF-MS/MS analysis (LC-MS/MS). Therefore, as expected, we have found that the Nano-LC-MS/MS analysis of the digested hapten-BSA glycoconjugate presented several advantages compared to the MALDI-MS/MS analysis of the glycosylated peptides. This was mainly due to the chromatographic separation, which restrains the ionization suppression effect present during the MALDI-MS analysis.^[36,37] Moreover, the CID-MS/MS analyses are low-energy compared with the high-energy CID-MS/MS experiments realized with the TOF/TOF-MS/MS instrument. For that reason, it is well known that the low energy CID-MS/MS analysis of peptides generally produces unique peptide diagnostic b- and y-ions, which provide a bidirectional analysis of their sequence.^[38]

Nano LC-MS/MS was carried out on both tryptic digests and GluC V8 digests. The results were submitted to the Mascot library, in order to identify the peptides belonging to the BSA. For the analysis of the tryptic digests, the Mascot library gave a match with two serum albumin protein isoforms from the *B. taurus* species: serum albumin precursor ([gi|1351907](#)) and serum albumin ([gi|74267962](#)).

4.3.4.1. CID-MS/MS of the tryptic digested peptides

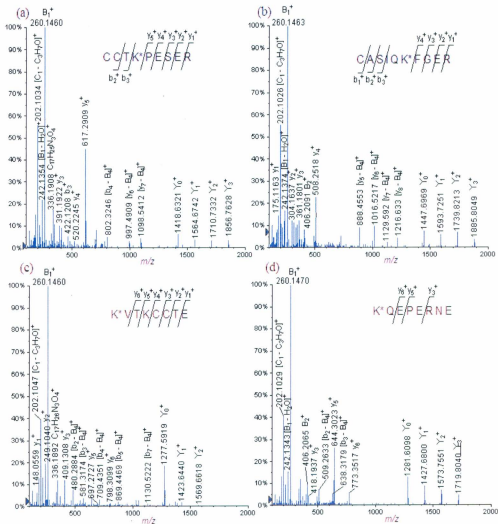
The detected peptides that matched to these two isoforms are reported in Table C.5. The detected peptides covered 58% of the sequence of the serum albumin protein from *B. taurus* ([gi|74267962](#)) and in addition, four unique peptides for this protein were identified [Table C.5]: TVMENFVAFVGK at m/z 671.3491 (+2), ETYGDMA²CCAK at m/z 710.7665 (+2), LGEYGFQNELIVR at m/z 769.4033 (+2) and MPCAEDYLSLILNR at m/z 847.9184 (+2) (charge of the molecular ions are shown in brackets). Similarly, the digested peptides covered 57% of the sequence of the serum albumin precursor from *B. taurus* ([gi|1351907](#)) and one unique peptide was detected for this protein: ETYGDMA²CCEKQEPER at m/z 706.6155 (+3). All the peptides that could not be assigned to any protein hits were presumed to be potentially glycosylated peptides.

The low-energy CID-MS/MS analysis of the extracted precursor ions extracted from the tetrasaccharide-spacer-BSA neoglycoconjugate and the corresponding glycopeptide product ions identified are shown in Table 4.1. Thus, the LC-MS/MS analyses permitted us to identify the carbohydrate occupancy located on the following 18 lysine glycation sites indicated as follows: Lys 140, Lys 155, Lys 156, Lys 204, Lys 211, Lys 228, Lys 235, Lys 304, Lys 374, Lys 401, Lys 420, Lys 437, Lys 455, Lys 463, Lys 495, Lys 498, Lys 547 and Lys 559. In this justification and for simplification purposes, we only decided to describe two examples of the low energy CID-MS/MS analyses of the glycopeptides. Figures 4.5(a) (Table C.6) and 4.5(b) exhibit two examples of the gas-phase fragmentation resulting from the CID-MS/MS analysis of the glycopeptides, as well as the product ion calculated masses and the mass deviation of the experimental

masses compared to the calculated ones. In these low-energy CID analyses, we observed the consecutive losses of carbohydrate moieties from the intact glycopeptides, as well as the formation of peptide product ions characterized by either the loss of the entire carbohydrate moiety and/or the partially fragmented carbohydrate product ions.

Table 4.1. Tryptic glycopeptides identified of the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate.

Precursor ion <i>m/z</i> (charge)	Mr(expt)	Mr(calc)	Deviation Da	Missed cleavage	Peptide
729.8826 (+2)	1457.7506	1457.7389	0.0117	1	KHK*P (Lys 559)
733.3170 (+2)	1464.6194	1464.7698	-0.1504	1	QIK*K (Lys 547)
770.3757 (+2)	1538.7369	1538.7338	0.0031	1	ADEK*K (Lys 155)
807.9138 (+2)	1613.8129	1613.7964	0.0165	0	K*FWGK (Lys 156)
828.0811 (+3)	2481.2213	2481.2005	0.0208	2	LKECCDK*P LLEK (Lys 304)
832.7770 (+3)	2495.3092	2495.3146	-0.0054	1	LK*HLVDEPQNLIK (Lys 401)
863.7965 (+3)	2588.3677	2588.3572	0.0105	0	K*VPQVSTPTLVEVSR (Lys 437)
883.9655 (+2)	1765.9164	1765.9085	0.0079	1	SLGK*VGTR (Lys 455)
922.0941 (+3)	2763.2605	2763.2493	0.0112	1	LAK*EYEATLEECCAK (Lys 374)
969.5086 (+2)	1937.0026	1936.9868	0.0158	1	TPVSEK*VTK (Lys 495)
970.4900 (+2)	1938.9738	1938.9772	-0.0034	1	EK*VLTSSAR (Lys 211)
976.0173 (+2)	1950.0201	1950.0085	0.0116	1	ALK*AWSVAR (Lys 235)
990.4710 (+3)	2968.3911	2968.3886	0.0025	2	LK*PDPNTLCDEFKADEK (Lys 140)
1073.0144 (+2)	2144.0143	2144.0082	0.0061	1	CASIQK*FGER (Lys 228)
1058.4652 (+2)	2114.9159	2114.9123	0.0036	1	CCTK*PESER (Lys 463)
1208.5857 (+2)	2415.1569	2415.1284	0.0285	1	VTK*CCTESLVNR (Lys 498)
1160.2172 (+3)	3477.6297	3477.6385	-0.0088	1	QNCDQFEK*LGEYGFQNALIVR (Lys 420)
1169.5978 (+2)	2337.1810	2337.1583	0.0227	1	GACLLPK*IETMR (Lys 204)



For example, the CID-MS/MS analysis of the glycopeptide precursor ion $[M + 2H]^{2+}$ at m/z 1058.4652 extracted from CCTK*PESER (Lys 463) [Figure 4.5(a), Table C.6], showed the following peptide ions which were assigned to be formed by the loss of the carbohydrate moieties: Y_3^+ at m/z 1856.7628, Y_2^+ at m/z 1710.7332, Y_1^+ at m/z 1564.6742 and Y_0^+ at m/z 1418.6321. As observed for the previous MALDI-MS/MS analysis of the glycosylated peptides, we were able to identify peptide fragment ions presenting a loss of a carbohydrate moiety: $[y_7 - B_4]^+$ at m/z 1098.5412, $[y_6 - B_4]^+$ at m/z 997.4909 and $[b_4 - B_4]^+$ at m/z 802.3246. The following carbohydrate-spacer product ions were also detected: B_2^+ at m/z 406.2071, $C_{17}H_{26}N_3O_4^+$ at m/z 336.1908, C_1^+ at m/z 276.1435, B_1^+ at m/z 260.1460, $[B_1 - H_2]^+$ at m/z 258.1313, $[B_1 - H_2O]^+$ at m/z 242.1354, $C_{11}H_{16}N_3O_2^+$ at m/z 222.1184, $[C_1 - C_3H_7O]^+$ at m/z 202.1034, $^{2,4}A_1^+$ at m/z 184.0928 and $[B_1 - C_5H_8O_2]^+$ at m/z 160.0945. The further continuous coverage of the peptide sequence was also achieved by detection of the y- and b-ions.

The CID-MS/MS analysis of the precursor ion $[M + 2H]^{2+}$ at m/z 1073.0144 isolated from the glycopeptide CASIQK*FGER (Lys 228), followed the same gas-phase fragmentation pathway as that of $[M + 2H]^{2+}$ precursor ion at m/z 1058.4652, described above [Figure 4.5(b)].

4.3.4.2. CID-MS/MS of the GluC V8 digested peptides

The Mascot library search of the product ions obtained by the LC-MS/MS analysis of the GluC V8 digests matches the same serum albumin proteins as observed for the analysis of the tryptic digests, namely, serum albumin precursor (*B. taurus*) ([gi|1351907](#)) and serum albumin (*B. taurus*) ([gi|74267962](#)). The sequences of the peptides which matched these two isoforms are displayed in Table C.7. The protein

sequence coverage was 42% for the serum albumin ([gi|74267962](#)) and 45% for the precursor serum albumin ([gi|1351907](#)), which is consistently lower than the coverage obtained with the tryptic digests.

The following 17 glycation sites were also confirmed by low-energy CID-MS/MS analyses: Lys 65, Lys 75, Lys 88, Lys 100, Lys 117, Lys 151, Lys 183, Lys 197, Lys 256, Lys 266, Lys 304, Lys 309, Lys 336, Lys 374, Lys 420, Lys 455 and Lys 495 (Table 4.2). To illustrate the fragmentation of the glycated peptides identified after the digestion of the hapten-BSA glycoconjugate with the endoproteinase GluC V8, two CID- spectra [Figures 4.5(c) and 4.5(d)] are displayed.

The LC-MS/MS analysis of the glycopeptide precursor ion $[M + 2H]^{2+}$ at m/z 987.9658 isolated from the glycopeptide K*VTKCCTE (Lys 495) [Figure 4.5(c), Table C.8], enabled us to observe peptide product ions obtained by loss of the fragmented tetrasaccharide portion of the glycoconjugate: Y_2^+ at m/z 1569.6618, Y_1^+ at m/z 1423.6440, Y_0^+ at m/z 1277.5919 and $[Y_0 - H_2O]^+$ at m/z 1259.5710. The y - and b -product ions permitted us to cover the peptide sequence, and the detection of fragmented peptide formed by loss of the tetrasaccharide moiety helped for the localization of the carbohydrate hapten on the Lys 495 residue: $[b_7 - B_4]^+$ at m/z 1130.5222, $[b_5 - B_4]^+$ at m/z 869.4469, $[b_4 - B_4]^+$ at m/z 709.4351, $[b_3 - B_4]^+$ at m/z 581.3174 and $[b_2 - B_4]^+$ at m/z 480.2884. In addition, we noted that the same fragmented carbohydrate product ions, which were detected for the trypsin digestion of the hapten-BSA glycoconjugate, were also present: B_2^+ at m/z 406.2084, $C_{17}H_{26}N_3O_4^+$ at m/z 336.1892, C_1^+ at m/z 276.1415, B_1^+ at m/z 260.1460, $[B_1 - H_2]^+$ at m/z 258.1307, $[B_1 - H_2O]^+$ at m/z 242.1380, $[C_1 - C_3H_7O]^+$ at m/z 202.1047, $^{2,4}A_1^+$ at m/z 184.0948 and $[B_1 - C_5H_8O_2]^+$ at m/z 160.0912.

Table 4.2. Glycopeptides identified in the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate digested with the endoprotease GluC V8.

Precursor ion <i>m/z</i> (charge)	Mr(expt)	Mr(calc)	Deviation Da	Missed cleavage	Peptide
698.3598 (+2)	1394.7051	1394.6803	0.0248	0	K*LGE (Lys 420)
790.3774 (+2)	1578.7403	1578.7288	0.0115	1	K*QEPE (Lys 117)
844.3721 (+2)	1686.7297	1686.7499	-0.0202	1	EFK*ADE (Lys 151)
894.4637 (+2)	1786.9128	1786.8976	0.0152	0	HVK*LVNE (Lys 65)
862.9539 (+2)	1723.8933	1723.8754	0.01179	1	VTK*LVTD (Lys 256)
897.4200 (+2)	1792.8255	1792.8176	0.0079	0	K*SHCIAE (Lys 309)
902.4801 (+2)	1802.9456	1802.9289	0.0167	0	LTkVHK*E (Lys 266)
962.1552 (+2)	2883.4439	2883.4093	0.0346	0	VSRSLGK*VGTRCCTKPE (Lys 455)
987.9658 (+2)	1973.9171	1973.8949	0.0222	0	K*VTKCCTE (Lys 495)
989.9697 (+2)	1977.9248	1977.9154	0.0094	2	K*QEPERNE (Lys 117)
992.4568 (+2)	1982.8989	1982.884	0.0149	1	CCDK*PLLE (Lys 304)
995.4702 (+2)	1988.9258	1988.8912	0.0346	1	FAK*TCVADE (Lys 75)
1013.0142 (+2)	2024.0138	2024.0123	0.0015	0	LCK*VASLRE (Lys 100)
1048.5225 (+2)	2095.0305	2094.9984	0.0321	1	K*SLHTLFGDE (Lys 88)
1097.0595 (+2)	2192.1045	2192.0909	0.0136	0	DK*GACLLPKIE (Lys 197)
1124.5018 (+2)	2246.9890	2246.9876	0.0014	1	DK*DVCKNYQE (Lys 336)
1336.1446 (+2)	2670.2746	2670.2728	0.0018	0	LLYYANK*YNGVFQE (Lys 183)
1156.132 (+2)	2310.2495	2310.2345	0.0150	0	YAVSVLLRLAK*E (Lys 374)

The other example that we selected to discuss was the glycopeptide K*QEPERNE (Lys 117) [Figure 4.5(d)]. The low-energy CID-MS/MS analysis of this glycopeptide precursor ion $[M + H]^{2+}$ at *m/z* 989.9697 extracted from K*QEPERNE (Lys 117) also exhibited the expected y- and b- product ions, which cover the peptide sequence, the carbohydrate product ions and also the diagnostic peptide product ions, which have lost the tetrasaccharide B₄ portion.

During the LC-MS/MS analysis of the digested hapten-BSA glycoconjugate with both trypsin and GluC V8, we were able to identify more glycation sites than we did

during the MALDI-TOF/TOF-MS/MS analysis of the digested glycoprotein. LC-MS/MS analysis allowed us to confidently identify 30 glycation sites on the lysine residues with total sequence coverage of 92%, whereas only 5 glycation sites were identified during the MALDI-MS/MS analysis of the digested glycoconjugate. The reason for this discrepancy seems to lie entirely on the formation of various glycoforms, as a result of the protocol for the carbohydrate conjugation which involves, intentionally, an insufficient amount of glycation agent to affect a complete derivatization of the protein carrier. Moreover, during the study we also observed identically located lysine residues in their derivatized and underivatized form. For example, the Lys 183 was detected through the underivatized doubly charged peptide ion LLYYANKYNGVFQE at m/z 861.4378 and also through its derivatized form on the doubly charged peptide ion LLYYANK*YNGVFQE at m/z 1336.1446 obtained from the GluC V8 digestion of the neoglycoconjugate. These observations support the fact that the sample is composed of a mixture of glycoforms.

The total glycation sites identified during the study are summarized on Figure 4.6(a). In addition, Figure 4.6(b) displays the 3D representation of the BSA obtained by homology modeling using the Swiss-Pdb Viewer software, where the glycated lysine residues are highlighted in red.^[39,40] We can observe that most of the glycated lysines are located on the outer-surface region of the protein. Thus, we can presume that the carbohydrate hapten functionalization and occupancy on the outer surface of the BSA will depend entirely on its tertiary structure.

(a)

```
1 MKWVTFISLL LFFSSAYSRG VFRDTHKSE IAHRFKDLGE EHFGLVLIA
51 FSQYLQCCPF DEHVK*LVNEL TEFAK*TCVAD ESHAGCEK*SL HTLFGDELCK*
101 VASLRETYGD MADCCA*KQEP ERNECFLSHK DDSPDLPKLK* PDPNTLCDEF
151 K*ADEK*K*FWGK YLYEIARRHP YFYAPELLYY ANK*YNGVFQE CCQAEDK*GAC
201 LLPK*IETMRE K*VLTSSARQR LRCASIQK*PG ERALK*AWSVA RLSQKFPKAE
251 FVEVTK*LVTD LTKVHK*ECCH GDLLCADDR ADLAKYICDN QDTISSKLKE
301 CCDK*P LLEK*S HCIAEVEKDA IPENLPPLTA DFAEDK*DVCK NYQEAKDAFL
351 GSFLEYEYRR HPEYAVSVLL RLA*K*EYEATL EECCAADDPH ACYSTVFDFKL
401 K*HLVDEPQNL IKQNCDFEK* LGEYGFQNEL IVRYTRK*VPQ VSTPTLVEVS
451 RSLGK*VGTRC CTK*PESERMP CAEDYLSLIL NRLCVLHEKT PVSEK*VTK*CC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIK*KQT
551 ALVELLKHK*P KATEEQLKTV MENFVAFVGK CCAADDKEAC FAVEGPKLVV
601 STQTALA
```

(b)

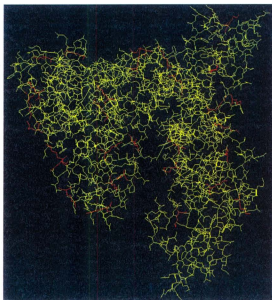


Figure 4.6. (a) BSA sequence where the glycation sites are indicated by an asterisk (red = identified on tryptic digests, blue = identified on GluC V8 digests and red and underlined = identified on both tryptic and GluC V8 digests) and (b) 3D structure of the BSA. The glycated lysine residues are highlighted in red (Swiss-Pdb Viewer software).

4.4. Conclusion

We have analyzed a *B. anthracis* neoglycoconjugate vaccine using the MALDI- and LC-ESI-tandem mass spectrometry. This was achieved to localize the carbohydrate hapten occupancies and their conjugation sites in the synthetic tetrasaccharide *B. anthracis* exosporium-squaric acid linker-BSA protein carrier neoglycoconjugate. The single stage MALDI-MS analysis of the glycoconjugate was carried out to verify the average hapten:BSA ratio, which was found to be 5.4:1.

The high-energy MALDI-MS/MS analyses were carried out on tryptic and GluC V8 digests of the conjugate, and these allowed us to sequence and reveal only five glycation sites of the identified glycopeptides. It is noteworthy that the MALDI-MS/MS experiments of the digested GluC V8 endoproteinase enabled the identification of three new glycation sites, which were not revealed during the analysis of the tryptic digests. The nano-LC-MS/MS analysis of the tryptic and GluC V8 digests permitted us to identify many more glycation sites (30 were identified). Fewer carbohydrate-spacer fragment ions were observed during the LC-MS/MS analysis when compared with the MALDI-MS/MS analyses, which are done with higher collision energies that enhance carbohydrate cleavages.

The use of GluC V8 endoproteinase for the neoglycoconjugate digestion enabled us to identify more carbohydrate hapten occupancies and their respective glycation sites than when only trypsin was used. Thus, a total of 18 glycation sites were identified in the CID-MS/MS analysis of the tryptic digest, and 12 novel glycation sites were observed through the LC-MS/MS analysis of the GluC V8 digest of the carbohydrate-BSA conjugate.

Finally, the central conclusion drawn from this work, as expected, is the nano-LC-MS/MS analysis of tryptic and GluC V8 digests is a reliable tool for the determination of carbohydrate occupancy and glycation sites of carbohydrate-protein conjugate vaccines.

Furthermore, as presented in this manuscript, we were able to use successfully ESI-CID-MS/MS to circumvent the complicated interpretation of the MALDI-CID fragmentation of the product ions containing the carbohydrate moieties. We are proposing to use in our future work, some more advanced activation methods for ESI-MS/MS analysis, such as electron transfer dissociation and electron capture dissociation measured by, FTICR-MS and Orbitrap mass spectrometers.^[42-44] These non-ergodic activation methods are known specifically to lead dissociation of the peptide backbone solely and to leave intact any further chemical modifications in the selected precursor ion.^[41,42]

Appendix C. Supplementary data

Supplementary data are associated with this chapter (page 248)

References

- [1] C. Galanos, O. Luderitz. In *Handbook of Endotoxins*, Chemistry of Endotoxins E. T. Rietschel (Ed). Elsevier: Amsterdam, The Netherlands, **1984**, 46.
- [2] E. T. Rietschel, H. W. Wollenweber, H. Brade, U. Zahring, B. Linder, U. Seydel, H. Bradaezek, G. Barnickel, H. Labischinski, P. Giesbrecht. In *Handbook of Endotoxins, Chemistry of Endotoxins*. E. T. Rietschel (Ed). Elsevier, Amsterdam, The Netherlands, **1984**, 187.
- [3] O. Holst, H. Brade. In *Bacterial Endotoxic Lipopolysaccharides*, Molecular Biochemistry and Cellular Biology. D. C. Morrison, J. L. Ryan (Eds). CRC Press, Boca Raton, Florida, USA, **1992**, 135.
- [4] C.R.H. Raetz. Biochemistry of Endotoxins. *Ann. Rev. Biochem.* **1990**, 59, 129.
- [5] J. Kubler-Kielb, E. Vinogradov, G. Ben-Menachem, V. Pozsgay, J. B. Robbins, R. Schneerson. Saccharide/protein conjugate vaccines for *Bordetella* species: Preparation of saccharide, development of new conjugation procedures, and physico-chemical and immunological characterization of the conjugates. *Vaccine* **2008**, 26, 3587.
- [6] V. Pozsgay. Recent Developments in Synthetic Oligosaccharide-Based Bacterial Vaccines. *Curr. Top. Med. Chem.* **2008**, 8, 126.
- [7] C. Grandjean, A. Boutonnier, B. Dassy, J. M. Fournier, L. A. Mulard. Investigation towards bivalent chemically defined glycoconjugate immunogens prepared from acid-detoxified lipopolysaccharide of *Vibrio cholerae* O1, serotype Inaba. *Glycoconj. J.* **2009**, 26, 41.
- [8] R. S. Daum, D. Hogerman, M. B. Rennels, K. Bewley, F. Malinoski, E. Rothstein, K. Reisinger, S. Block, H. Keyserling, M. Steinhoff. Infant immunization with

- pneumococcal CRM₁₉₇ vaccines: effect of saccharide size on immunogenicity and interactions with simultaneously administrated vaccines. *J. Infect. Dis.* **1997**, 176, 445.
- [9] D. J. Lefeber, J. P. Kamerling, J. F. G. Vliegthart. Synthesis of Streptococcus pneumoniae Type 3 Neoglycoproteins Varying in oligosaccharide chain length, loading and carrier protein *Chem. Eur. J.* **2001**, 7, 4411.
- [10] L. C. Paoletti, D. L. Kasper, F. Michon, J. DiFabio, H. J. Jennings, T. D. Tosteson, M. R. Wessels. Effects of chain length on the immunogenicity in rabbits of group B Streptococcus type III oligosaccharide-tetanus toxoid conjugates. *J. Clin. Invest.* **1992**, 89, 203.
- [11] A. Chernyak, S. Kondo, T. K. Wade, M. D. Meeks, P. M. Alzari, J. M. Fournier, R. K. Taylor, P. Kováč, W. F. Wade. Induction of protective immunity by synthetic Vibrio cholerae hexasaccharide derived from *V. cholerae* O1 Ogawa lipopolysaccharide bound to a protein carrier. *J. Infect. Dis.* **2002**, 185, 950.
- [12] F. Jahouh, R. Saksena, D. Aiello, A. Napoli, G. Sindona, P. Kováč, J. H. Banoub. Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by Matrix-Assisted Laser Desorption/Ionization tandem mass spectrometry. *J. Mass Spectrom.* **2010**, 45, 1148.
- [13] M. Mock, A. Fouet. Anthrax. *Annu. Rev. Microbiol.* **2001**, 55, 647.
- [14] F. G. Priest. In *Bacillus subtilis and other Gram-positive bacteria: Biochemistry, Physiology, and Molecular Biology*. A. L. Sonenshein, J. A. Hoch, R. Losick (Eds). American Society for Microbiology, Washington, D. C., USA, **1993**, 3.

- [15] W. L. Nicholson, N. Munakata, G. Horneck, H. J. Melosh, P. Setlow. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 548.
- [16] I. Boutiba-Ben Boubaker, S. Ben Redjeb. *Bacillus anthracis*: causative agent of anthrax. *Tunis Med.* **2001**, *79*, 642.
- [17] D. D. Williams, O. Benedek, C. L. Jr. Turnbough. Species-specific peptide ligands for the detection of *Bacillus anthracis* spores. *Appl. Environ. Microbiol.* **2003**, *69*, 6288.
- [18] D. J. Chabot, A. Scorpio, S. A. Tobery, S. F. Little, S. L. Norris, A. M. Friedlander. Anthrax capsule vaccine protects against experimental infection. *Vaccine* **2004**, *23*, 43.
- [19] J. M. Daubenspeck, H. Zeng, P. Chen, S. Dong, C. T. Steichen, N. R. Krishna, D. G. Pritchard, Jr. C. L. Turnbough. Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis* exosporium. *J. Biol. Chem.* **2004**, *279*, 30945.
- [20] R. Adamo, R. Saksena, P. Kováč. Studies towards a conjugate vaccine for Anthrax: synthesis of the tetrasaccharide side chain of the *Bacillus anthracis* exosporium. *Helv. Chim. Acta* **2006**, *89*, 1075.
- [21] R. Saksena, R. Adamo, P. Kováč. Immunogens related to the synthetic tetrasaccharide side chain of the *Bacillus anthracis* exosporium. *Bioorg. Med. Chem.* **2007**, *15*, 4283.
- [22] A. Dell, H. R. Morris. Glycoprotein structure determination by mass spectrometry. *Science* **2001**, *291*, 2351.
- [23] J. Hui, H. Desaire, V. Y. Butnev, G. R. Bousfield. Glycoprotein profiling by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 750.

- [24] N. D. Rawlings, A. J. Barrett. Families of serine peptidases. *Meth. Enzymol.* **1994**, 244, 19.
- [25] G. R. Drapeau, Y. Boily, J. Houmard. Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J. Biol. Chem.* **1972**, 247, 6720.
- [26] J. J. Birktoft, K. Breddam. In Proteolytic enzymes: Glutamyl endopeptidases. *Methods Enzymol.* **1994**, 244, 114.
- [27] A. Chernyak, A. Karavanov, Y. Ogawa, P. Kováč. Conjugating oligosaccharides to proteins by squaric acid diester chemistry; rapid monitoring of the progress of conjugation, and recovery of the unused ligand. *Carbohydr. Res.* **2001**, 330, 479.
- [28] A. L. Yergey, J. R. Coorsen, P. S. Backlund Jr., P. S. Blank, G. A. Humphrey, J. Zimmerberg. De novo sequencing of peptides using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **2002**, 13, 784.
- [29] B. Domon, C. Costello. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, 5, 397.
- [30] P. Roepstorff, J. Fohlman. Letter to the editors. *Biol. Mass Spectrom.* **1984**, 11, 601.
- [31] R. S. Johnson, S. A. Martin, K. Biemann, J. T. Stults, J. T. Watson. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, 59, 2621.
- [32] T. J. Cornish, R. J. Cotter. A curved field reflectron time-of-flight mass spectrometer for the simultaneous focusing of metastable product ions. *Rapid Commun. Mass Spectrom.* **1994**, 8, 781.

- [33] R. Kaufmann, D. Kirsch, B. Spengler. Sequencing of peptides in a time-of-flight mass spectrometer: evaluation of post source decay following Matrix-Assisted Laser Desorption Ionization (MALDI). *Int. J. Mass Spectrom. Ion Process.* **1994**, 131, 355.
- [34] T. Satoh, T. Sato, A. Kubo, J. Tamura. Tandem Time-of-Flight mass spectrometer with high precursor ion selectivity employing spiral ion trajectory and improved offset parabolic reflectron. *J. Am. Soc. Mass Spectrom.* **2011**, 22, 797.
- [35] I. A. Papayannopoulos. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **1995**, 14, 49.
- [36] W. I. Burkitt, A. E. Giannakopoulos, F. Sideridou, S. Bashir, P. J. Derrick. Discrimination effects in MALDI-MS of mixtures of peptides-analysis of the proteome. *Austr. J. Chem.* **2003**, 56, 369.
- [37] R. Kratzer, C. Eckerskorn, M. Karas, F. Lottspeich. Suppression effects in enzymatic peptide ladder sequencing using ultraviolet - matrix assisted laser desorption/ionization - mass spectrometry. *Electrophoresis* **1998**, 19, 1910.
- [38] A. Schlosser, W. D. Lehmann. Five-membered ring formation in unimolecular reactions of peptides: a key structural element controlling low-energy collision-induced dissociation of peptides. *J. Mass Spectrom.* **2000**, 35, 1382.
- [39] K. Arnold, L. Bordoli, J. Kopp, T. Schwede. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, 22, 195.
- [40] M. C. Peitsch. Protein modeling by E-mail. *Biol. Technol.* **1995**, 13, 658.
- [41] R. A. Zubarev, N.L. Kelleher, F.W. McLafferty. ECD of multiply charged protein cations. A non-ergodic process, *J. Am. Chem. Soc.* **1998**, 120, 3265.

- [42] A. Stensballe, O. N. Jensen, J. V. Olsen, K. F. Haselmann, R. A. Zubarev. Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1793.
- [43] E. Mirgorodskaya, P. Roepstorff, R. A. Zubarev. Localization of *O*-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* **1999**, *71*, 4431.
- [44] K. Hakansson, H. J. Cooper, M. R. Emmett, C. E. Costello, A. G. Marshall, C. L. Nilsson. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an *N*-glycosylated tryptic peptic to yield complementary sequence information. *Anal. Chem.* **2001**, *73*, 4530.

CHAPTER 5: Revealing the glycation sites in neoglycoconjugate models formed by conjugation of the antigenic monosaccharide hapten of *Vibrio cholerae* O1 serotype Ogawa with the BSA protein carrier using LC-ESI-QqTOF-MS/MS

This chapter has been submitted: Farid Jahouh,¹ Rina Saksena,² Pavol Kováč² and Joseph H. Banoub^{1,3*} **2011**, *J. Mass Spectrom.*, submitted.

¹Department of Chemistry, Memorial University of Newfoundland, Saint John's, NL, A1B 3V6, Canada

²NIDDK, LBC, National Institutes of Health, Bethesda, MD 20892-0815, USA

³Department of Fisheries and Oceans Canada, Science Branch, Special Projects, Saint John's, NL, Canada, A1C 5X1

Abstract

We present the determination of glycation sites in synthetic neoglycoconjugate vaccine models formed by conjugation of the antigenic monosaccharide hapten of *Vibrio cholerae* O1 serotype Ogawa to BSA carrier using nano-LC-ESI-QqTOF-MS/MS. MALDI-TOF/TOF-MS/MS analyses of tryptic digests of the glycoconjugates having hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1 revealed only three glycation sites (lysine residues: Lys 235, Lys 437 and Lys 455). Trypsin and GluC V8 proteases were used for the digestion of the neoglycoconjugates. The digests gave complementary structural information and maximized the number of recognized glycation sites. We report identification of 20, 27 and 33 glycation sites using LC-ESI-QqTOF-MS/MS analysis of a series of digested synthetic neoglycoconjugate vaccines with a hapten:BSA ratio of respectively 4.3:1, 6.6:1 and 13.2:1. We also propose that all the glycated lysine residues are located mainly on the outer surface of the protein.

5.1 Introduction

Different studies have been carried out on the analysis of glycoproteins using mass spectrometry.^[1-3] Novel soft ionization techniques such as ESI and MALDI were applied successfully for the determination of the glycation sites of *N*- and *O*-glycoproteins, as well as for the general carbohydrate structure determination.^[3,4] Generally, two methods have been applied. The first involves cleavage of the glycans from the protein by hydrolysis, followed by a purification step and MS analysis of the intact carbohydrate portions.^[5,6] It has been described that sometimes it is necessary to derivatize the released polysaccharides in order to obtain a sufficiently high *m/z* signal of the released carbohydrate.^[7] The second approach is based on the digestion of the entire glycoprotein with endoproteases, followed by MS and MS/MS analyses of the digests obtained.^[8,9] The main advantage of this method is that the product ions formed correspond to the glycopeptide fragments, which contain information about the exact glycation site.^[10]

In our previous work, we have reported the determination of the glycation sites of a series of neoglycoconjugate models by MALDI-TOF-MS/MS. These glycoconjugates were composed of the spacer-equipped terminal monosaccharide antigen of the *O*-PS of *V. cholerae* O1, serotype Ogawa covalently attached to the protein carrier BSA.^[11] We were able to determine the hapten:BSA ratios for the different synthetic glycoconjugate preparations as being 4.3:1, 6.6:1 and 13.2:1. The obtained glycoconjugates were digested with trypsin and the resulting digests were analyzed by MALDI-TOF/TOF-MS/MS in order to determine the conjugation sites between the spacer-equipped carbohydrate and the carrier protein. In that study, we reported that three glycation sites

on the Lys 235, Lys 437 and Lys 455 residues were identified on the three analyzed glycoconjugates. We observed that digestion of our hapten-BSA glycoconjugates with the protease trypsin may have not been best choice. Additionally, it is well known that ion suppression effects can occur during the MALDI-TOF-MS experiments, resulting in a curtailed MALDI-MS spectrum.^[12,13] This suppression phenomenon may be explained by competition in the ionization of the different analytes and by presence of contaminants in the sample.

We recently reported the determination of the glycation sites in an experimental *Bacillus anthracis* neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-MS/MS.^[14] The MALDI-TOF-MS analysis of the intact glycoconjugate permitted us to determine the hapten-to-BSA ratio which was found to be 5.4:1. The carbohydrate-protein vaccine was digested using two different proteases: trypsin and the GluC V8 endoproteinase. The different digests were subsequently subjected to MALDI-TOF/TOF-MS/MS and LC-ESI-QqTOF-MS/MS analysis. The study revealed that the LC-ESI-QqTOF-MS/MS analysis of the different digests allowed identification of 30 glycation sites, than 5 shown by the MALDI-TOF/TOF-MS/MS analysis of the same digests. We thus concluded that the LC-ESI-QqTOF-MS/MS analysis of both tryptic and GluC V8 digests was the more efficient method for the investigation of the glycation sites in synthetic carbohydrate-protein glycoconjugates.

In the present work, we applied the same strategy to test *V. cholerae* O1 serotype Ogawa neoglycoconjugate vaccines.^[14] We digested the synthetic glycoconjugate *V. cholerae* O1 serotype Ogawa, formed by the conjugation of the monosaccharide antigen to the BSA with different hapten-to-BSA ratios (4.3:1, 6.6:1 and 13.2:1), separately with

trypsin and the GluC V8 endoproteinase. The different digests were then analyzed by LC-ESI-QqTOF-MS/MS and the glycopeptides were sequenced manually in order to determine the glycation sites.

5.2. Material and Methods

5.2.1. Preparation of the hapten-BSA neoglycoconjugate

The synthesis of the hapten-BSA glycoconjugates has been described previously.^[15] The synthesis started with preparation of the methyl 6-hydroxyhexanoyl α -glycoside of the upstream terminal moiety of the O-PS of *V. cholerae* serogroup O1 serotype Ogawa antigen [4-(3-deoxy-L-glycero)-2-O-methyl- α -D-perosamine]. The spacer-equipped sugar was then treated with 1,2-diaminoethane which converted the methyl ester into the corresponding amino amide. Reaction of the latter with diethyl squarate produced a squarate monoester. The hapten thus obtained was conjugated with BSA (Sigma Aldrich, Saint Louis, MO, USA) as described.^[15] Isolation of the neoglycoconjugates was carried out by ultrafiltration using centrifugal filters with a molecular weight cut-off of 30,000 Da.

5.2.2. Neoglycoconjugates digestion with different proteases

The digestions of the hapten-BSA glycoconjugates were carried out with trypsin and Glu-C V8 protease (Sigma Aldrich, Saint Louis, MO, USA). Thus, 100 μ g of the glycoconjugate was dissolved in a mixture of 0.1% RapiGest SF Surfactant (1 μ g, Waters, USA) in 50 mM of NH_4HCO_3 (100 μ L) at a pH of 8.0 and reduced by treatment with 2 μ L of 10 mM of DTT (Sigma Aldrich, Saint Louis, MO, USA) for 30 min at room temperature, followed by alkylation with 2 μ L of a 50 mM iodoacetamide (Sigma

Aldrich, Saint Louis, MO, USA) for 1 h at room temperature. A portion (50 μg) of the BSA glycoconjugate was digested with trypsin using 20 ng/mL of trypsin dissolved in NH_4HCO_3 (50 mM, 1 mL, pH 7.8) at a trypsin:glycoprotein ratio of 1:25 (w/w) and incubated at 37 °C overnight with shaking. The other 50 μg of the BSA glycoconjugate was digested using the GluC V8 endoprotease dissolved in NH_4HCO_3 (50 mM, 1 mL, pH 7.8) at a protease:glycoprotein ratio of 1:25 (w/w) and incubated at 37 °C overnight with shaking. The sample was then dried under vacuum and the residue was dissolved in 20 μL of 1% acetic acid (Sigma-Aldrich, Oakville, ON, Canada). An aliquot of each sample (10 μL) was then cleaned up using ZipTip C18 (Millipore, Bedford, MA, USA) before mass spectral analysis.

5.2.3. LC-ESI-QqTOF-CID-MS/MS analysis

The peptides were separated on a DIONEX UltiMate3000 Nano LC System (Germering, Germany). 250 fmol of digested glycoprotein was dissolved in 0.1% TFA and loaded onto a precolumn (300 μm i.d. \times 5 mm, C18 PepMap100, 5 μm (LC Packing, Sunnyvale, CA)) in order to desalt and concentrate the sample. After their elution from the precolumn, the mixture of peptides and glycopeptides was separated on a nanoflow analytical column (75 μm i.d. \times 15 cm, C18 PepMap 100, 3 μm , 100 Å, (LC Packing, Sunnyvale, CA)) at a flow rate of 180 nL/min. The mobile phase eluents used were composed of 0.1% FA/0.01% TFA/2% ACN (A) and 0.08% FA/0.008% TFA/98% ACN (B). The elution started with 0% B for 10 min, followed by a gradient of 0-60% B in 55 min and 60-90% B in 3 min and was kept at 90% B for 3 min. The MS/MS analysis of the eluted peptides and glycopeptides was accomplished using an Applied Biosystems API-QSTAR XL quadrupole orthogonal time-of-flight (QqTOF)-MS/MS hybrid tandem

mass spectrometer (Applied Biosystems International-MDS Sciex, Foster City, CA, USA) equipped with a nano-electrospray source (Protana XYZ manipulator) which produces the electrospray through a PicoTip needle (10 μm i.d., New Objectives, Woburn, MA, USA) carrying a voltage of 2400 V. The TOF analyzer was calibrated using a renin solution (1 pmol/ μL) and looking for the ions at m/z 586.9815 and m/z 879.9723. The collision energies used during the CID-MS/MS analyses were determined automatically using the Information Dependent Acquisition method integrated in the Analyst software.

5.3. Results

The intent of this study was to determine the glycation sites in synthetic hapten-BSA glycoconjugates prepared by conjugation of the antigenic monosaccharide hapten of *V. cholerae* O1 serotype Ogawa to BSA having hapten-to-BSA ratios of 4.3:1, 6.6:1 and 13.2:1.

The nomenclature of carbohydrate fragments which will be used to identify the various glycopeptides was described by Domon and Costello^[3], whereas the nomenclature used to identify the true peptide fragments is the one described by Roepstorff *et al.*^[16] and recently modified by Johnson and coworkers.^[17]

The Nano-LC-ESI-QqTOF-MS/MS analysis of the tryptic digests of the neoglycoconjugates with a hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1 allowed the identification of two serum albumin protein isoforms from the *Bos taurus* species on the Mascot library: serum albumin precursor ([gi|1351907](#)) with the following sequence coverage: 61% for the neoglycoconjugate possessing a hapten:BSA ratio of 4.3:1, 61 % for the neoglycoconjugate with a hapten:BSA ratio of 6.6:1 and 47 % for the

neoglycoconjugate with a hapten:BSA ratio of 13.2:1. In addition, during the Mascot library search, we also identified the serum albumin protein ([gi|74267962](#)) for the analysis of all the neoglycoconjugates.

The submission of the LC-ESI-QqTOF-MS/MS data of the GluC V8 digests to the Mascot library allowed identification of two serum albumin protein isoforms from the *B. taurus* species: serum albumin precursor ([gi|1351907](#), sequence coverage of 55 % for the neoglycoconjugate with a hapten:BSA ratio 4.3:1, 45 % for the neoglycoconjugate with a hapten:BSA ratio 6.6:1 and 40 % for the neoglycoconjugate with a hapten:BSA ratio 13.2:1) and serum albumin ([gi|74267962](#)).

We have identified molecular masses of glycopeptides produced from the digested glycoconjugates by comparison of the m/z values of the obtained molecular ions which did not correspond to BSA digests in the Mascot report, with the calculated m/z values of all the possible glycopeptides. Thus, the masses of the glycopeptides were identified by the addition of either one or two carbohydrate-hapten residues (513.23 Da), corresponding to the m/z values of the formed glycopeptides. In other words, the m/z values of the peptides which matched to the theoretical mass of a “peptide + carbohydrate-linker” were identified, extracted and subjected to low-energy CID-MS/MS for sequencing to identify the glycation sites.

5.3.1. Mass spectrometry determination of the glycation sites on the neoglycoconjugate possessing a hapten:BSA ratio of 4.3:1

The previously described strategy allowed us to identify 11 glycation sites through the sequencing of the following glycopeptides obtained from the tryptic digests of the neoglycoconjugate with a hapten:BSA ratio of 4.3:1 (Table 5.1): SLGK*VGTR

(Lys 455) at m/z 665.8810 (+2), EK*VLTSSAR (Lys 211) at m/z 752.4233 (+2), ALK*AWSVAR (Lys 235) at m/z 757.9261 (+2), K*QTALVELLK (Lys 548) at m/z 828.4992 (+2), CCTK*PESER (Lys 463) at m/z 840.3949 (+2), LK*PDPNTLCDEFKADEK (Lys 140) at m/z 845.0908 (+3), DTHK*SEIAHR (Lys 28) at m/z 853.9445 (+2), GACLLPK*IETMR (Lys 204) at m/z 951.5091 (+2), K*VPQVSTPTLVEVSR (Lys 437) at m/z 1077.1129 (+2), LAK*EYEATLEECCA (Lys 374) at m/z 1164.5375 (+2) and QNCDQFEK*LGEYGFQNALIVR (Lys 420) at m/z 1014.8462 (+3).

Table 5.1. Tryptic glycopeptides identified during LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate with a hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1.

Peptide		Missed cleavage	Hapten:BSA ratio 4.3:1		Hapten:BSA ratio 6.6:1		Hapten:BSA ratio 13.2:1	
Sequence (asterisk = glycation site)	Calculated <i>m/z</i> (charge)		Observed <i>m/z</i> (charge)	Deviation (Da)	Observed <i>m/z</i> (charge)	Deviation (Da)	Observed <i>m/z</i> (charge)	Deviation (Da)
SLGK*VGTR (Lys 455)	665.8643 (+2)	1	665.8810 (+2)	0.0168	665.8694 (+2)	0.0052	665.8699 (+2)	0.0057
EK*VLTSSAR (Lys 211)	752.3987 (+2)	1	752.4233 (+2)	0.0246	752.4114 (+2)	0.0127	752.3971 (+2)	-0.0016
ALK*AWSVAR (Lys 235)	757.9143 (+2)	1	757.9261 (+2)	0.0118	757.9189 (+2)	0.0046	757.9248 (+2)	0.0105
K*QTALVELLK (Lys 548)	828.4770 (+2)	1	828.4992 (+2)	0.0222	828.4872 (+2)	0.0103	828.4857 (+2)	0.0087
CCTK*PESER (Lys 463)	840.3662 (+2)	1	840.3949 (+2)	0.0287	840.3788 (+2)	0.0126	840.3769 (+2)	0.0107
LK*PDPNTLCDEFKADEK (Lys 140)	845.0720 (+3)	3	845.0908 (+3)	0.0188	845.0839 (+3)	0.0119	845.0911 (+3)	0.0191
DTHK*SEIAHR (Lys 28)	853.9209 (+2)	1	853.9445 (+2)	0.0236	853.9366 (+2)	0.0158	-	-
GACLLPK*IETMR (Lys 204)	951.4892 (+2)	1	951.5091 (+2)	0.0199	951.5014 (+2)	0.0122	951.5006 (+2)	0.0114
K*VPQVSTPTLVEVSR (Lys 437)	1077.0887 (+2)	0	1077.1129 (+2)	0.0243	1077.0985 (+2)	0.0099	1077.0975 (+2)	0.0089
LAK*EYEATLEECCAK (Lys 374)	1164.5347 (+2)	2	1164.5375 (+2)	0.0028	1164.5613 (+2)	0.0266	1164.5540 (+2)	0.0193
QNCDQFEK*LGEYGFQNALIVR (Lys 420)	1014.8220 (+3)	1	1014.8462 (+3)	0.0242	1014.8390 (+3)	0.0170	1014.8534 (+3)	0.0314
HKPK*ATEEQLK (Lys 561)	607.9779 (+3)	3	-	-	607.9966 (+3)	0.0187	608.0033 (+3)	0.0254

LSQK*FPK (Lys 245)	680.8716 (+2)	3	-	-	680.8722 (+2)	0.0007	680.8755 (+2)	0.0040
LCVLHEK*TPVSEK (Lys 489)	685.0222 (+3)	2	-	-	685.0256 (+3)	0.0034	685.0310 (+3)	0.0088
CASIQK*FGER (Lys 228)	854.9142 (+2)	1	-	-	854.9284 (+2)	0.0143	854.9290 (+2)	0.0149
VTK*CCTESLVNR (Lys 498)	990.4743 (+2)	1	-	-	990.4813 (+2)	0.0071	990.4930 (+2)	0.0188
FK*DLGEEHFK (Lys 36)	881.9304 (+2)	2	-	-	-	-	881.9491 (+2)	0.0188
LKECCDK*PULEK (Lys 304)	1023.5103 (+2)	3	-	-	-	-	1023.5348 (+2)	0.0245

Figure 5.1(a) displays the mass spectra obtained from the extraction of the precursor ion at m/z 752.4233 (+2) corresponding to the glycosylated peptide EK*VLTSSAR (Lys 211). The CID-MS/MS analysis of this glycosylated peptide reveals a signature of the carbohydrate moiety through the following product ions: B^+ at m/z 262.1265, $[B - H_2O]^+$ at m/z 244.1142, $[B - 2H_2O]^+$ at m/z 226.1047, $[^{2,5}A - 2H_2O]^+$ at m/z 152.0663, $[B - C_4H_6O_3]^+$ at m/z 142.0815 and $[^{2,5}A - CH_4O_3]^+$ at m/z 124.0708 (Figure 5.2). Moreover, we were also able to detect the entire peptide attached to the linker portion: Y^+ at m/z 1242.6794 and Z^+ at m/z 1224.6785. Also, we noticed that the product ion at m/z 336.1918 could be assigned to two different structures: $[B + C_4H_{10}O]^+$ ($C_{15}H_{30}NO_7^+$) and $C_{17}H_{26}N_3O_4^+$ (Figure 5.2). However, since the product at m/z 336.1918 was also detected in our previous work on *Bacillus anthracis* neoglycoconjugate vaccine model, it was assigned to the structure: $[C_{17}H_{26}N_3O_4]^+$ corresponding to a linker portion attached to a lysine residue (Figure 5.2).^[14] Similarly, the product ion $[C_{11}H_{16}N_3O_2]^+$ at m/z 222.1187 was produced from a different linker portion attached to a lysine residue. This latter product ion was also detected in our previous work on the glycation sites determination of *Bacillus anthracis* neoglycoconjugate vaccine model.^[14]

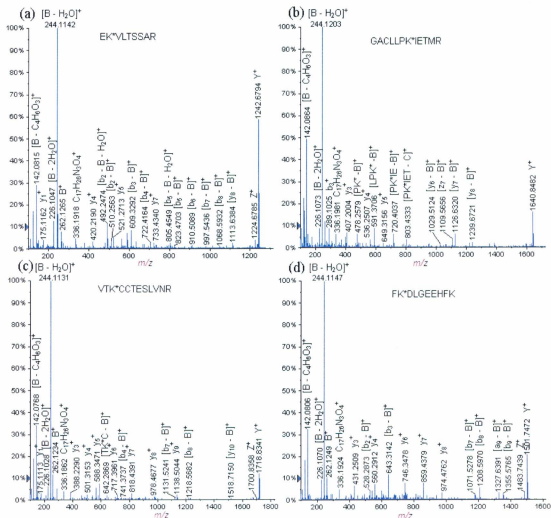


Figure 5.1. LC-ESI-QqTOF-MS/MS spectra of the tryptic glycopeptides (a) EK*VLTSSAR (Lys 211) at m/z 752.4233 (+2), (b) GACLLPK*1ETMR (Lys 204) at m/z 951.5091 (+2), (c) VTK*CTTESLVNR (Lys 498) at m/z 990.4813 (+2) and FK*DLGEEHFK (Lys 36) at m/z 881.9491 (+2).

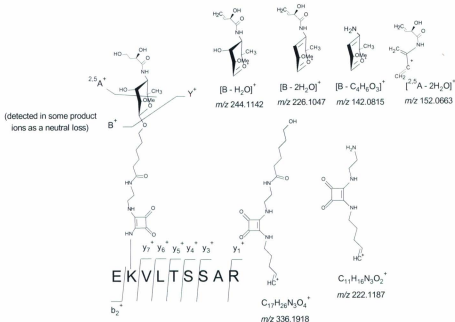


Figure 5.2. Different product ions involving the fragmentation of the carbohydrate hapten observed during the LC-ESI-QqTOF-MS/MS analysis of the tryptic glycopeptide EK*VLTTSSAR (Lys 211) at m/z 752.4233 (+2).

The CID-MS/MS analysis of this glycosylated peptide EK*VLTTSSAR (Lys 211) at m/z 752.4233 (+2), allowed us to observe the formation of product ions corresponding to glycopeptide fragments obtained by the loss of their carbohydrate portion: $[y_8 - B]^+$ at m/z 1113.6384, $[b_8 - B]^+$ at m/z 1068.5932, $[b_7 - B]^+$ at m/z 997.5436, $[b_6 - B]^+$ at m/z 910.5089, $[b_5 - B]^+$ at m/z 823.4703, $[b_5 - B - H_2O]^+$ at m/z 805.4549, $[b_4 - B]^+$ at m/z 722.4164, $[b_4 - B - H_2O]^+$ at m/z 704.4081, $[b_3 - B]^+$ at m/z 609.3292, $[b_3 - B - H_2O]^+$ at m/z 591.3171, $[b_2 - B]^+$ at m/z 510.2563 and $[b_2 - B - H_2O]^+$ at m/z 492.2474. The series

of detected y- and b-ions which represented the sequence of the glycopeptides are listed in Table 5.2.

Table 5.2. LC-ESI-QqTOF-MS/MS analysis of the glycosylated peptide EK*VLTSSAR (Lys 211) at m/z 752.4233 (+2).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
Y^+	1242.6683	1242.6794	0.0111
Z^+	1224.6577	1224.6785	0.0208
$[y_8 - B]^+$	1113.6252	1113.6384	0.0132
$[b_8 - B]^+$	1068.5572	1068.5932	0.0360
$[b_7 - B]^+$	997.5201	997.5436	0.0235
$[b_6 - B]^+$	910.4880	910.5089	0.0209
$[b_5 - B]^+$	823.4560	823.4703	0.0143
$[b_5 - B - H_2O]^+$	805.4454	805.4549	0.0095
y_7^+	733.4197	733.4340	0.0143
$[b_4 - B]^+$	722.4083	722.4164	0.0081
$[b_4 - B - H_2O]^+$	704.3977	704.4081	0.0104
y_6^+	634.3513	634.3576	0.0063
$[b_3 - B]^+$	609.3243	609.3292	0.0049
$[b_3 - B - H_2O]^+$	591.3137	591.3171	0.0034
y_5^+	521.2673	521.2713	0.0040
$[b_2 - B]^+$	510.2558	510.2563	0.0005
$[b_2 - B - H_2O]^+$	492.2452	492.2474	0.0022
y_4^+	420.2196	420.2190	-0.0006
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1918	0.0000
y_3^+	333.1875	333.1851	-0.0024
B^+	262.1290	262.1265	-0.0025
$[B - H_2O]^+$	244.1179	244.1142	-0.0037
$[B - 2H_2O]^+$	226.1079	226.1047	-0.0032
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1187	-0.0050
y_1^+	175.1184	175.1162	-0.0022
$[^{25}A - 2H_2O]^+$	152.0706	152.0663	-0.0043
$[B - C_4H_8O_3]^+$	142.0863	142.0815	-0.0048
$[^{25}A - CH_4O_3]^+$	124.0757	124.0708	-0.0049

Figure 5.1(b) shows another example CID-MS/MS analysis of a glycated peptide detected during the analysis of the tryptic digests of the hapten-BSA neoglycoconjugate with a hapten:BSA ratio of 4.3:1: GACLLPK**IETMR* (Lys 204) at m/z 951.5091 (+2).

The LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests of the neoglycoconjugate with a hapten:BSA ratio of 4.3:1 permitted the localization of 10 glycation sites through the sequencing of the following glycopeptides (Table 5.3): K**SHCIAE* (Lys 309) at m/z 679.3320 (+2), LTKVHK**E* (Lys 266) at m/z 684.3867 (+2), K**VTKCCTE* (Lys 495) at m/z 769.8677 (+2), K**QEPERNE* (Lys 117) at m/z 771.8852 (+2), K**SLHTLFGDE* (Lys 88) at m/z 830.4932 (+2), DKGACLLPK**IE* (Lys 204) at m/z 878.9783 (+2), AK**DAFLGSFLYE* (Lys 346) at m/z 937.4871 (+2), KK**FWGKYLYE* (Lys 156) at m/z 937.9877 (+2), LLYYANK**YNGVFQE* (Lys 183) at m/z 1118.0711 (+2) and VSRSLGK**VGTRCCTKPE* (Lys 455) at m/z 816.7646 (+3).

Table 5.3. GluC V8 glycopeptide digests identified during LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate with a hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1.

Peptide		Missed cleavage	Hapten:BSA ratio 4.3:1		Hapten:BSA ratio 6.6:1		Hapten:BSA ratio 13.2:1	
Sequence (asterisk = glycation site)	Calculated m/z (charge)		Observed m/z (charge)	Deviation (Da)	Observed m/z (charge)	Deviation (Da)	Observed m/z (charge)	Deviation (Da)
K*SHCIAE (Lys 309)	679.3189 (+2)	0	679.3320 (+2)	0.0131	679.3343 (+2)	0.0154	679.3284 (+2)	0.0095
LTKVHK*E (Lys 266)	684.3745 (+2)	0	684.3867 (+2)	0.0122	684.3851 (+2)	0.0106	684.3795 (+2)	0.0050
K*VTKCCTE (Lys 495)	769.8575 (+2)	0	769.8677 (+2)	0.0102	769.8720 (+2)	0.0145	769.8668 (+2)	0.0093
K*QEPERNE (Lys 117)	771.8678 (+2)	2	771.8852 (+2)	0.0174	771.8795 (+2)	0.0117	771.8883 (+2)	0.0205
K*SLHTLFGDE (Lys 88)	830.4093 (+2)	1	830.4932 (+2)	0.0839	830.4246 (+2)	0.0153	830.4196 (+2)	0.0103
DKGACLLPK*IE (Lys 204)	878.9556 (+2)	0	878.9783 (+2)	0.0227	878.9669 (+2)	0.0113	878.9643 (+2)	0.0087
AK*DAFLGSFLYE (Lys 346)	937.4590 (+2)	1	937.4871 (+2)	0.0281	937.4759 (+2)	0.0169	937.4798 (+2)	0.0208
KK*FWGKYLVE (Lys 156)	937.9824 (+2)	0	937.9877 (+2)	0.0053	937.9862 (+2)	0.0038	937.9648 (+2)	-0.0176
LLYYANK*YNGVFQE (Lys 183)	1118.0465 (+2)	0	1118.0711 (+2)	0.0246	1118.0587 (+2)	0.0122	1118.0511 (+2)	0.0046
VSRSLGK*VGTRCCTKPE (Lys 455)	816.7456 (+3)	0	816.7646 (+3)	0.0190	816.7583 (+3)	0.0127	816.7483 (+3)	0.0027
K*TPVSE (Lys 489)	587.2979 (+2)	0	-	-	587.3103 (+2)	0.0124	587.3041 (+2)	0.0062
GPK*LVVSTQTALA (Lys 597)	899.4959 (+2)	0	-	-	899.5080 (+2)	0.0121	899.5092 (+2)	0.0133

LLKHKPK*ATEE (Lys 561)	903.9960 (+2)	1	-	-	904.0190 (+2)	0.0230	903.9916 (+2)	-0.0044
K*QIKKQTALVE (Lys 544)	900.0117 (+2)	0	-	-	900.0234 (+2)	0.0117	900.0165 (+2)	0.0048
VTK*LVTD (Lys 256)	644.8478 (+2)	0	-	-	-	-	644.8498 (+2)	0.0020
HVK*LVNE (Lys 65)	676.3588 (+2)	0	-	-	-	-	676.3647 (+2)	0.0059
VEK*DAIPE (Lys 318)	707.3534 (+2)	2	-	-	-	-	707.3578 (+2)	0.0044
CCDK*PLLE (Lys 304)	774.3520 (+2)	1	-	-	-	-	774.3662 (+2)	0.0142
LCK*VASLRE (Lys 100)	794.9162 (+2)	0	-	-	-	-	794.9163 (+2)	0.0001
VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463)	987.8230 (+3)	0	-	-	-	-	987.8313 (+3)	0.0083
KK*FWGK*YLYE (Lys 156, 160)	1194.5985 (+2)	0	-	-	-	-	1194.5626 (+2)	-0.0359
NFVAFVDK*CCAADDKE (Lys 580)	1201.5300 (+2)	3	-	-	-	-	1201.5350 (+2)	0.0050

Figure 5.3 displays two examples of glycopeptides identified during the LC-ESI-Qq-TOF-MS/MS analysis of the GluC V8 digests of the neoglycoconjugate with a hapten:BSA ratio of 4.3:1: (a) K*SHCIAE (Lys 309) at m/z 679.3320 (+2) and (b) VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7646 (+3). The fragmentation pathway followed by the glycopeptide VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7646 (+3) allowed us to identify the glycation site exclusively on the Lys 455 residue (Figure 5.3(b), Table 5.4). The precursor ion VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7646 (+3) produced characteristic carbohydrate moiety product ions, identified as follows: B^+ at m/z 262.1247, $[B - H_2O]^+$ at m/z 244.1136, $[B - 2H_2O]^+$ at m/z 226.1042, $[^{2,5}A - 2H_2O]^+$ at m/z 152.0669, $[B - C_4H_6O_3]^+$ at m/z 142.0824 and $[^{2,5}A_1 - CH_4O_3]^+$ at m/z 124.0714. Furthermore, we detected the following product ion Y^{2+} at m/z 1094.0368 corresponding to the entire glycopeptide minus the loss of the entire carbohydrate portion. In addition we also observed product ions corresponding to distinct fragmentation of the peptide backbone minus the loss of the entire carbohydrate moiety; needless to say that these product ions included the previously derivatized lysine: $[y_{12} - B]^+$ at m/z 1644.7886, $[y_{11} - B]^+$ at m/z 1587.7061, $[b_9 - B]^+$ at m/z 1136.5123, $[b_8 - B]^+$ at m/z 1079.5651, $[b_{16} - B]^{2+}$ at m/z 1020.5277, $[b_7 - B]^+$ at m/z 980.5699, $[b_{15} - B]^{2+}$ at m/z 972.0032, $[b_{14} - B]^{2+}$ at m/z 907.9123 and $[GK^* - B]^+$ at m/z 438.2288. Therefore, these product ions allowed us to sequence and characterize this glycopeptide correctly.

Finally, the following product ions, characteristic of a spacer portion attached to a lysine residue were also identified as follows: $[C_{17}H_{26}N_3O_4]^+$ at m/z 336.1787 and $[C_{11}H_{16}N_3O_2]^+$ at m/z 222.1242. The bidirectional sequence of the glycosylated peptide was also determined through the identification of the y- and b-ions (Figure 5.3(b), Table 5.4).

Table 5.4. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7646 (+3).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
$[y_{12} - B]^+$	1644.7821	1644.7886	-0.0065
$[y_{11} - B]^+$	1587.7607	1587.7061	0.0546
$[b_9 - B]^+$	1136.5422	1136.5123	0.0299
y_9^+	1108.4868	1108.4700	0.0168
Y^{2+}	1094.0541	1094.0368	0.0172
$[b_8 - B]^+$	1079.6208	1079.5651	0.0557
y_8^+	1051.4654	1051.4701	-0.0047
$[b_{16} - B]^{2+}$	1020.5278	1020.5277	0.0001
$[b_7 - B]^+$	980.5524	980.5699	-0.0175
$[b_{15} - B]^{2+}$	972.0014	972.0032	-0.0018
y_7^+	950.4177	950.4802	-0.0625
$[b_{14} - B]^{2+}$	907.9539	907.9123	0.0416
y_6^+	794.3166	794.3096	0.0070
y_5^+	634.2859	634.2643	0.0216
b_6^+	600.3469	600.3362	0.0107
b_5^+	543.3255	543.3074	0.0181
y_4^+	474.2553	474.2656	-0.0103
$[GK^* - B]^+$	438.2342	438.2288	0.0054
b_4^+	430.2414	430.2307	0.0107
y_3^+	373.2076	373.2188	-0.0112
b_3^+	343.2094	343.2028	0.0066
$C_{17}H_{26}N_7O_4^+$	336.1918	336.1787	0.0131
B^+	262.1290	262.1247	0.0043
y_2^+	245.1126	245.1127	-0.0001
$[B - H_2O]^+$	244.1179	244.1136	0.0043
$[B - 2H_2O]^+$	226.1079	226.1042	0.0037
$C_{11}H_{16}N_7O_2^+$	222.1237	222.1242	-0.0005
b_2^+	187.1083	187.1204	-0.0121
$[^{2.5}A - 2H_2O]^+$	152.0706	152.0669	0.0037
$[y_1 - H_2]^+$	146.0453	146.0706	-0.0253
$[B - C_4H_6O_3]^+$	142.0863	142.0824	0.0039
$[^{2.5}A_1 - CH_4O_3]^+$	124.0757	124.0714	0.0043

To sum it up, a total of 20 glycation sites were identified during the LC-ESI-QqTOF-MS/MS analysis of the tryptic and GluC V8 digests of the neoglycoconjugate with a hapten:BSA ratio of 4.3:1 [Figure 5.4(a)] for total sequence coverage of 89%. Figure 5.5(a) displays the 3-D structure of the BSA represented using the Swiss-Pdb Viewer software, where the glycated lysine residues are highlighted in red.^[18,19] A quick glance at Figure 5.5(a) allowed us to observe that the glycated lysine residues are located on the outer surface of the protein.

- (a) 1 MKWVTFISLL LLFSSAYSRG VFRDTHK*SE IAHRFKDLGE EHFGLVLIA
51 FSQYLQCCPF DEHVKLVEL TEFAKTCVAD ESHAGCEK*SL HTLFGDELCK
101 VASLRETYGD MADCCAK*QEP ERNECFLSHK DSDPLPKLK* PDPNTLCDEF
151 KADEKK*FWGK YLYEIARRHP YFYAPPELLY ANK*YNGVFQE CCQAEKDGAC
201 LLPK*IETMRE K*VLTSSARQR LRCASIQK*FG ERALK*AWSVA RLSQK*FPKAE
251 FVEVTKLVTD LTKVHK*ECCH GDLLCADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEK*S HCIAAEVKDA IPENLPPLTA DFAEDKDVCN NYQEAK*DAFL
351 GSFLYEYSRR HPEYAVSVLL RLAK*EYEATL EECCAADDPH ACYSTVFDFKL
401 KHLVDEPQNL IKQNCDFEK* LGEYGFQNEL IVRYTRK*VPQ VSTPTLVEVS
451 RSLGK*VGTRC CTK*PESERMP CAEDYLSLIL NRLCVLHEK*TVSEK*VTCKC
501 TESLVNRRPC FSALTDPETY VPFAFDEKLF TFHADICTLP DTEKQIKK*QT
551 ALVELLKHHP KATEEQLKTV MENFVAFVGK CCAADDKEAC FAVEGPKLVV
601 STQTALA
- (b) 1 MKWVTFISLL LLFSSAYSRG VFRDTHK*SE IAHRFKDLGE EHFGLVLIA
51 FSQYLQCCPF DEHVKLVEL TEFAKTCVAD ESHAGCEK*SL HTLFGDELCK
101 VASLRETYGD MADCCAK*QEP ERNECFLSHK DSDPLPKLK* PDPNTLCDEF
151 KADEKK*FWGK YLYEIARRHP YFYAPPELLY ANK*YNGVFQE CCQAEKDGAC
201 LLPK*IETMRE K*VLTSSARQR LRCASIQK*FG ERALK*AWSVA RLSQK*FPKAE
251 FVEVTKLVTD LTKVHK*ECCH GDLLCADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEK*S HCIAAEVKDA IPENLPPLTA DFAEDKDVCN NYQEAK*DAFL
351 GSFLYEYSRR HPEYAVSVLL RLAK*EYEATL EECCAADDPH ACYSTVFDFKL
401 KHLVDEPQNL IKQNCDFEK* LGEYGFQNEL IVRYTRK*VPQ VSTPTLVEVS
451 RSLGK*VGTRC CTK*PESERMP CAEDYLSLIL NRLCVLHEK*TVSEK*VTCKC
501 TESLVNRRPC FSALTDPETY VPFAFDEKLF TFHADICTLP DTEKQIKK*QT
551 ALVELLKHHP K*ATEEQLKTV MENFVAFVGK CCAADDKEAC FAVEGPKLVV
601 STQTALA
- (c) 1 MKWVTFISLL LLFSSAYSRG VFRDTHKSE IAHRFK*DLGE EHFGLVLIA
51 FSQYLQCCPF DEHV*LVNEL TEFAKTCVAD ESHAGCEK*SL HTLFGDELCK*
101 VASLRETYGD MADCCAK*QEP ERNECFLSHK DSDPLPKLK* PDPNTLCDEF
151 KADEKK*FWGK* YLYEIARRHP YFYAPPELLY ANK*YNGVFQE CCQAEKDGAC
201 LLPK*IETMRE K*VLTSSARQR LRCASIQK*FG ERALK*AWSVA RLSQK*FPKAE
251 FVEVTK*LVTD LTKVHK*ECCH GDLLCADDR ADLAKYICDN QDTISSKLKE
301 CCDK*P*LEK*S HCIAAEVK*DA IPENLPPLTA DFAEDKDVCN NYQEAK*DAFL
351 GSFLYEYSRR HPEYAVSVLL RLAK*EYEATL EECCAADDPH ACYSTVFDFKL
401 KHLVDEPQNL IKQNCDFEK* LGEYGFQNEL IVRYTRK*VPQ VSTPTLVEVS
451 RSLGK*VGTRC CTK*PESERMP CAEDYLSLIL NRLCVLHEK*TVSEK*VTCKC
501 TESLVNRRPC FSALTDPETY VPFAFDEKLF TFHADICTLP DTEKQIKK*QT
551 ALVELLKHHP K*ATEEQLKTV MENFVAFVGK* CCAADDKEAC FAVEGPKLVV
601 STQTALA

Figure 5.4. Neoglycoconjugate model sequences where the glycation sites are indicated by an asterisk (red = identified on tryptic digests, blue = identified on GluC V8 digests and red and underlined = identified on both tryptic and GluC V8 digests) for the neoglycoconjugates with a hpaten-BSA ratio of (a) 4.3:1, (b) 6.6:1 and (c) 13.2:1.

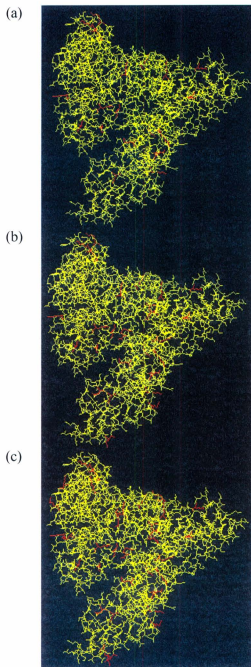


Figure 5.5. 3-D structure of the neoglycoconjugate model vaccines. The glycated lysine residues are highlighted in red (Swiss-Pdb Viewer software) for the neoglycoconjugates with a hapten:BSA ratio of (a) 4.3:1, (b) 6.6:1 and (c) 13.2:1.

5.3.2. Mass spectrometry determination of the glycation sites on the neoglycoconjugate possessing a hapten:BSA ratio of 6.6:1

The LC-ESI-QqTOF-MS/MS analysis of the tryptic digests of the neoglycoconjugate with a hapten:BSA ratio of 6.6:1 allowed identification of 16 glycation sites on the following glycopeptides (Table 5.1): SLGK*VGTR (Lys 455) at m/z 665.8694 (+2), EK*VLTSAR (Lys 211) at m/z 752.4114 (+2), ALK*AWSVAR (Lys 235) at m/z 757.9189 (+2), K*QTALVELLK (Lys 548) at m/z 828.4872 (+2), CCTK*PESER (Lys 463) at m/z 840.3788 (+2), LK*PDPNTLCDEFKADEK (Lys 140) at m/z 845.0839 (+3), DTHK*SEIAHR (Lys 28) at m/z 853.9366 (+2), GACLLPK*IETMR (Lys 204) at m/z 951.5014 (+2), K*VPQVSTPTLVEVSR (Lys 437) at m/z 1077.0985 (+2), LAK*EYEATLEECCA (Lys 374) at m/z 1164.5613 (+2), QNCQFEK*LGEYGFQNALIVR (Lys 420) at m/z 1014.8390 (+3), HKPK*ATEEQLK (Lys 561) at m/z 607.9966 (+3), LSQK*FPK (Lys 245) at m/z 680.8722 (+2), LCVLHEK*TPVSEK (Lys 489) at m/z 685.0256 (+3), CASIQK*FGER (Lys 228) at m/z 854.9284 (+2) and VTK*CCTESLVNR (Lys 498) at m/z 990.4813 (+2).

Figure 5.1(c) displays an example of a glycopeptide which was identified during the CID-MS/MS analysis of the tryptic digests of the neoglycoconjugate with a hapten:BSA ratio of 6.6:1: VTK*CCTESLVNR (Lys 498) at m/z 990.4813 (+2).

The neoglycoconjugate with a hapten:BSA ratio of 6.6:1 was also digested with the GluC V8 endoproteinase and subsequently analyzed by LC-ESI-QqTOF-MS/MS. The following glycation sites were discovered in this way after manual sequencing of the peptides which were suspected to be glycation (same mass than a "peptide + carbohydrate linker") and these are listed in Table 5.3: K*SHCIAE (Lys 309) at m/z 679.3343 (+2),

LTKVHK*E (Lys 266) at m/z 684.3851 (+2), K*VTKCCTE (Lys 495) at m/z 769.8720 (+2), K*QEPERNE (Lys 117) at m/z 771.8795 (+2), K*SLHTLFGDE (Lys 88) at m/z 830.4246 (+2), DKGACLLPK*IE (Lys 204) at m/z 878.9669 (+2), AK*DAFLGSFLYE (Lys 346) at m/z 937.4759 (+2), KK*FWGKLYLE (Lys 156) at m/z 937.9862 (+2), LLYYANK*YNGVFQE (Lys 183) at m/z 1118.0587 (+2), VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7583 (+3), K*TPVSE (Lys 489) at m/z 587.3103 (+2), GPK*LVVSTQTALA (Lys 597) at m/z 899.5080 (+2), LLKHKPK*ATEE (Lys 561) at m/z 904.0190 (+2) and K*QIKKQTALVE (Lys 544) at m/z 900.0234 (+2).

The LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests of the neoglycoconjugate with a hapten:BSA ratio of 6.6:1 allowed us to identify 14 glycation sites. Moreover, the combination of the identified glycation sites from both the tryptic and GluC V8 digests gives us a total of 27 glycation sites with total sequence coverage of 90 % [Figure 5.4(b)].

In addition, Figure 5.5(b) which shows the 3-D structure of the BSA, with the glycated lysine residues highlighted in red, enabled us to conclude that the glycated lysine residues are located on the outer surface of the BSA.

5.3.3. Mass spectrometry determination of the glycation sites on the neoglycoconjugate possessing a hapten:BSA ratio of 13.2:1

The neoglycoconjugate with a hapten:BSA ratio of 13.2:1 was also digested with trypsin and LC-ESI-Qq-TOF-MS/MS analysis of the digests allowed the identification of the following glycopeptides (Table 5.1): SLGK*VGTR (Lys 455) at m/z 665.8699 (+2), EK*VLTSSAR (Lys 211) at m/z 752.3971 (+2), ALK*AWSVAR (Lys 235) at m/z 757.9248 (+2), K*QTALVELLK (Lys 548) at m/z 828.4857 (+2), CCTK*PESER (Lys

463) at m/z 840.3769 (+2), LK*PDPNTLCDEFKADK (Lys 140) at m/z 845.0911 (+3), GACLLPK*IETMR (Lys 204) at m/z 951.5006 (+2), K*VPQVSTPTLVEVSR (Lys 437) at m/z 1077.0975 (+2), LAK*EYEATLEECCAK (Lys 374) at m/z 1164.5540 (+2), QNCQFEK*LGEYGFQNALIVR (Lys 420) at m/z 1014.8534 (+3), HKPK*ATEEQLK (Lys 561) at m/z 608.0033 (+3), LSQK*FPK (Lys 245) at m/z 680.8755 (+2), LCVLHEK*TPVSEK (Lys 489) at m/z 685.0310 (+3), CASIQK*FGER (Lys 228) at m/z 854.9290 (+2), VTK*CTESLVNR (Lys 498) at m/z 990.4930 (+2), FK*DLGEEHFK (Lys 36) at m/z 881.9491 (+2) and LKECCDK*PLLEK (Lys 304) at m/z 1023.5348 (+2). Total of 17 glycation sites were thus discovered. Figure 5.1(d) an example of identified glycopeptides, which were only discovered on the neoglycoconjugate with a hapten:BSA ratio of 13.2:1: FK*DLGEEHFK (Lys 36) at m/z 881.9491 (+2).

Finally, the analysis by LC-ESI-QqTOF-MS/MS of the GluC V8 digests permitted identification of the following glycopeptides (Table 5.3): K*SHCIAE (Lys 309) at m/z 679.3284 (+2), LTKVHK*E (Lys 266) at m/z 684.3795 (+2), K*VTKCCTE (Lys 495) at m/z 769.8668 (+2), K*QEPERNE (Lys 117) at m/z 771.8883 (+2), K*SLHTLFGDE (Lys 88) at m/z 830.4196 (+2), DKGACLLPK*IE (Lys 204) at m/z 878.9643 (+2), AK*DAFLGSFLYE (Lys 346) at m/z 937.4798 (+2), KK*FWGKLYLE (Lys 156) at m/z 937.9648 (+2), LLYYANK*YNGVFQE (Lys 183) at m/z 1118.0511 (+2), VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7483 (+3), K*TPVSE (Lys 489) at m/z 587.3041 (+2), GPK*LVVSTQTALA (Lys 597) at m/z 899.5092 (+2), LLKHKPK*ATEE (Lys 561) at m/z 903.9916 (+2), K*QIKKQTALVE (Lys 544) at m/z 900.0165 (+2), VTK*LVTD (Lys 256) at m/z 644.8498 (+2), HVK*LVNE (Lys 65) at m/z 676.3647 (+2), VEK*DAIPE (Lys 318) at m/z 707.3578 (+2), CCDK*PLLE (Lys

304) at m/z 774.3662 (+2), LCK*VASLRE (Lys 100) at m/z 794.9163 (+2) and NFVAFVDK*CCAADDKE (Lys 580) at m/z 1201.5350 (+2). Moreover, we also observed glycopeptides carrying two carbohydrate-linker residues: VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 987.8313 (+3) and KK*FWGK*YLYE (Lys 156, 160) at m/z 1194.5626 (+2).

It is important to note in this connection, that we have identified glycopeptides which contained two diagnostic glycation sites. Unquestionably, two glycopeptides with the following peptide sequence VSRSLGKVGTRCCTKPE were detected containing one glycation site on Lys 455, VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 816.7483 (+3), and also with two glycation sites on each of Lys 455 and Lys 463 residues: VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 987.8313 (+3). This observation supports the fact that the sample was composed of glycoforms. Figure 5.3(d) displays the spectra obtained during the CID-MS/MS analysis of the glycated peptide VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 987.8313 (+3), while the observed product ions and their mass-to-charge ratios are listed in Table 5.5.

In the product ion scan of the precursor ion at m/z 987.8313 (+3), we observed the product ion Y^{2+} at m/z 1350.6416, which was formed by loss of the carbohydrate moiety from one of the glycated lysine residues. The product ion $[Y - B]^{2+}$ at m/z 1220.0900 was initiated by the loss of the carbohydrate portions of the two glycated lysine residues. It is important to mention that the following series of diagnostic product ions: B^+ at m/z 262.1259, $[B - H_2O]^+$ at m/z 244.1156, $[B - 2H_2O]^+$ at m/z 226.1058, $[^{2-5}A - 2H_2O]^+$ at m/z 152.0673, $[B - C_4H_6O_3]^+$ at m/z 142.0837, $[^{2-5}A - CH_4O_3]^+$ at m/z 124.0733, $[C_{17}H_{26}N_3O_4]^+$ at m/z 336.1890 and $[C_{11}H_{16}N_3O_2]^+$ at m/z 222.1170 were formed

specifically from both the carbohydrate portions and the spacer moieties of the glycopeptide precursor ion. Furthermore, we have observed “proper” peptide product ions during this CID-MS/MS analysis, which were identified as follows: $[b_{14} - B]^+$ at m/z 1814.8000, $[b_{11} - B]^+$ at m/z 1393.8366, $[y_9 - B]^+$ at m/z 1360.5399, $[b_{10} - B]^+$ at m/z 1237.6267, $[y_7 - B]^+$ at m/z 1202.6109, $[b_9 - B]^+$ at m/z 1136.6248, $[b_8 - B]^+$ at m/z 1079.5776, $[b_7 - B]^+$ at m/z 980.5371, $[b_{14} - B]^{2+}$ at m/z 907.9442, $[b_{12} - B]^{2+}$ at m/z 777.4165, $[b_{11} - B]^{2+}$ at m/z 697.3653, $[CTK^* - B]^+$ at m/z 642.2561, $[y_3 - B]^+$ at m/z 625.3205, $[K^*VG - B]^+$ at m/z 537.2682, $[TK^* - B]^+$ at m/z 482.2529, $[GK^* - B]^+$ at m/z 438.2285 and $[K^* - B]^+$ at m/z 381.1820. All of these product ions were formed by the loss of the carbohydrate moiety exclusively, which allowed us to localize unambiguously the two different glycation sites.

Thus the CID-MS/MS sequencing of the previously mentioned glycopeptides allowed us to identify 22 glycation sites. Moreover, the combination of the CID-MS/MS analysis of the tryptic and GluC V8 digests enabled us to discover a total of 33 glycation sites with total sequence coverage of 88% [Figure 5.4(c)]. As previously observed for the neoglycoconjugates with hapten:BSA ratios of 4.3:1 and 6.6:1, the glycated lysine residues for the hapten-BSA glycoconjugate with a ratio of 13.2:1 are positioned on the outer surface of the protein [Figure 5.5(c)].

Table 5.5. LC-ESI-QqTOF-MS/MS analysis of the glycosylated peptide VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 987.8313 (+3).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
$[b_{14} - B]^+$	1814.9000	1814.8000	0.1000
$[b_{11} - B]^+$	1393.7910	1393.8366	-0.0456
$[y_9 - B]^+$	1360.5973	1360.5399	0.0574
Y^{2+}	1350.6700	1350.6416	0.0284
$[b_{10} - B]^+$	1237.6899	1237.6267	0.0632
$[Y - B]^{2+}$	1220.1091	1220.0900	0.0191
$[y_7 - B]^+$	1202.5282	1202.6109	-0.0827
$[b_9 - B]^+$	1136.6422	1136.6248	0.0174
$[b_8 - B]^+$	1079.6208	1079.5776	0.0432
$[b_7 - B]^+$	980.5524	980.5371	0.0153
$[b_{14} - B]^{2+}$	907.9539	907.9442	0.0097
$[b_{12} - B]^{2+}$	777.4148	777.4165	-0.0018
$[b_{11} - B]^{2+}$	697.3994	697.3653	0.0341
$[CTK^* - B]^+$	642.2910	642.2561	0.0349
$[y_3 - B]^+$	625.3181	625.3205	-0.0024
b_6^+	600.3469	600.3531	-0.0062
b_5^+	543.3255	543.3561	-0.0306
$[K^*VG - B]^+$	537.3026	537.2682	0.0344
$[TK^* - B]^+$	482.2604	482.2529	0.0075
$[GK^* - B]^+$	438.2342	438.2285	0.0057
b_4^+	430.2414	430.2185	0.0229
$[K^* - B]^+$	381.2127	381.1820	0.0307
b_1^+	343.2094	343.2214	-0.0120
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1890	0.0028
B^+	262.1290	262.1259	0.0031
$[B - H_2O]^+$	244.1179	244.1156	0.0023
$[B - 2H_2O]^+$	226.1079	226.1058	0.0021
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1170	0.0067
b_2^+	187.1083	187.1118	-0.0035
$[^{2,5}A - 2H_2O]^+$	152.0706	152.0673	0.0033
$[y_1 - H_2]^+$	146.0453	146.0758	-0.0305
$[B - C_4H_8O_3]^+$	142.0863	142.0837	0.0026
$[^{2,5}A - CH_4O_3]^+$	124.0757	124.0733	0.0024

5.4. Conclusion

The LC-ESI-QqTOF-MS/MS analysis of the tryptic and GluC V8 digests allowed identification of a higher number of glycation sites than previously reported for the neoglycoconjugates with a hapten:BSA ratio of 4.3:1, 6.6:1 and 13.2:1.^[11] Therefore, the present approach can be considered as a refinement of the previous analysis. As expected, the number of glycation sites in the neoglycoconjugates increases with the hapten:BSA ratio. Indeed, 20 glycation sites were observed for the neoglycoconjugate with a hapten:BSA ratio of 4.3:1, 27 for the neoglycoconjugate with a hapten:BSA ratio of 6.6:1 and 33 for the neoglycoconjugate with a hapten:BSA ratio of 13.2:1. Moreover, it was found that the glycated lysine residues are situated on the outer surface of the BSA.

In a future work, we plan to apply this method to find glycation sites on a synthetic neoglycoconjugate models formed by the conjugation of an antigenic hexasaccharide hapten of *V. cholerae* O1 serotype Ogawa and tetanus toxin protein carrier.

References

- [1] C. E. Costello, J. M. Contado-Miller, J. F. Cipollo. A glycomics platform for the analysis of permethylated oligosaccharide alditols. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1799.
- [2] B. Domon, C. E. Costello. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **1988**, *27*, 1534.
- [3] B. Domon, C. E. Costello. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, *5*, 397.
- [4] S. A. Carr, M. J. Huddleston, M. F. Bean. Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci.* **1993**, *2*, 183.
- [5] J. C. Rouse, J. E. Vath. On-the-probe sample cleanup strategies for glycoprotein-released carbohydrates prior to matrix-assisted laser desorption-ionization time-of-flight spectrometry. *Anal. Biochem.* **1996**, *238*, 82.
- [6] C. N. Scanlan, R. Pantophlet, M. R. Wormald, E. O. Saphire, R. Stanfield, I. A. Wilson, H. Katinger, R. A. Dwek, P. M. Rudd, D. R. Burton. The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of $\alpha\rightarrow 2$ mannose residues on the outer face of gp120. *J. Virol.* **2002**, *76*, 7306.
- [7] D. J. Harvey. N-(2-Diethylamino)Ethyl-4-Aminobenzamide derivative for high sensitivity mass spectrometric detection and structure determination of N-linked carbohydrates. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 862.

- [8] Y. Mechref, M. V. Novotny. Structural investigations of glycoconjugates at high sensitivity. *Chem. Rev.* **2002**, *102*, 321.
- [9] M. J. Huddleston, M. F. Bean, S. A. Carr. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal. Chem.* **1993**, *65*, 877.
- [10] A. Dell, H. R. Morris. Glycoprotein structure determination by mass spectrometry. *Science* **2001**, *291*, 2351.
- [11] F. Jahouh, R. Saksena, D. Aiello, A. Napoli, G. Sindona, P. Kováč, J. H. Banoub. Determination of the glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry. *J. Mass Spectrom.* **2010**, *45*, 1148.
- [12] M. Karas, R. Krüger. Ion formation in MALDI: the cluster ionization mechanism. *Chem. Rev.*, **2003**, *103*, 427.
- [13] R. Knochenmuss, A. Vertes. Time-delayed 2-pulse studies of MALDI matrix ionization mechanisms. *J. Phys. Chem. B*, **2000**, *104*, 5406.
- [14] F. Jahouh, S. J. Hou, P. Kováč, J. H. Banoub. Determination of the glycation sites of *Bacillus anthracis* neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-tandem mass spectrometry. *J. Mass Spectrom.* **2011**, *46*, 993.
- [15] R. Saksena, A. Chernyak, A. Karavanov, P. Kováč. Conjugating low molecular mass carbohydrates to proteins. 1. Monitoring the progress of conjugation. *Meth. Enzymol.* **2003**, *362*, 125.

- [16] P. Roepstorff, J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biol. Mass Spectrom.* **1984**, *11*, 601.
- [17] R. S. Johnson, S. A. Martin, K. Biemann, J. T. Stults, J. T. Watson. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, *59*, 2621.
- [18] K. Arnold, L. Bordoli, J. Kopp, T. Schwede. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195.
- [19] M. C. Peitsch. Protein modeling by E-mail. *Biol. Technol.* **1995**, *13*, 658.

CHAPTER 6: Determination of glycation sites by tandem mass spectrometry in a synthetic lactose-BSA conjugate, a vaccine model prepared by dialkyl squarate chemistry

This chapter has been published: Farid Jahouh,¹ Shu-jie Hou,² Pavol Kováč² and Joseph H. Banoub^{1,3*} **2011**, *Rapid Commun. Mass Spectrom.*, *accepted and in press*.

¹Department of Chemistry, Memorial University of Newfoundland, Saint John's, NL, A1B 3V6, Canada

²NIDDK, LBC, National Institutes of Health, Bethesda, MD 20892-0815, USA

³Department of Fisheries and Oceans Canada, Science Branch, Special Projects, Saint John's, NL, Canada, A1C 5X1

Abstract

Covalent attachment of lactose to BSA with a hapten:BSA ratio of 20:1 was prepared by a known method. The carbohydrate antigen (hapten)-BSA ratio in two different lactose-BSA glycoconjugate vaccine models obtained at different reaction times was determined by MALDI-TOF/TOF-MS to be 5.1:1 and 19.0:1. The glycation sites in the neoglycoconjugates were determined using nano-LC-ESI-QqTOF-MS/MS analysis of the trypsin and GluC V8 digests of the conjugates. Analysis of the digests gave complementary structural information and maximized the number of recognized glycation sites. For the neoglycoconjugate vaccine with a hapten:BSA ratio of 5.1:1, a total of 15 glycation sites were identified all located on the BSA lysine residues. The nano-LC-ESI-QqTOF-MS/MS of the tryptic and GluC V8 digests of the hapten-BSA glycoconjugate with a hapten:BSA ratio 19.0:1 allowed the identification of 30 glycation sites. Considering the average hapten:BSA ratio found in the neoglycoconjugate and the number of glycation sites identified, we concluded that the glycoconjugation results in formation of various glycoforms.

6.1. Introduction

Synthesis of carbohydrate-protein neoglycoconjugates has become an important avenue towards treatment of infectious diseases, as exemplified by the synthesis of carbohydrate based vaccines.^[1-4] Different methods have been applied for the synthesis of neoglycoconjugate vaccines. One of the most efficient methods for the synthesis of glycoconjugates is provided by the squaric acid chemistry, which allows single point attachment of the antigen to the protein carrier; it avoids the formation of cross-linking reactions, and allows recovery of the non-reacted squaric acid derivative (usually used in excess at the onset of the conjugation).^[5] Kamath *et al.* were first to monitor the conjugation using MALDI-TOF-MS.^[6] The first step in the conjugation process is the synthesis of the hapten in the form of a glycoside whose aglycone contains a primary amino group. The latter can react at pH 7.0 with a squaric acid diester at pH 7.0 to produce the corresponding amide ester. Subsequent reaction of the monoester with amine groups in the carrier protein at pH 9.0 forms the carbohydrate-protein glycoconjugate. More recently, Kováč's group developed a strategy that enables a one-pot preparation of a series of neoglycoconjugates with pre-determined carbohydrate:protein ratios.^[7-9] They were then able to make conjugates from fragments of the O-PS of *Vibrio cholerae* O1 serotype Ogawa^[9,10], and of Inaba^[11,12] with BSA to investigate the immunogenicity of these hapten-BSA glycoconjugates.

While the hapten:protein ratio has been generally measured using MALDI-TOF or SELDI-TOF mass spectrometry,^[10,13] the localization of the carbohydrate occupancies and the sites of conjugation has been more difficult to achieve.

We reported our first attempt to localize the conjugation sites in the hapten-BSA glycoconjugate made from the monosaccharide antigen of the *O*-PS of *V. cholerae* serotype Ogawa.^[14] The glycoconjugate was first digested with trypsin and the suspected glycosylated peptides were then analyzed by MALDI-TOF/TOF-MS/MS. However, presence of only three glycation sites were identified, despite the average molecular mass of the conjugate evidenced that it contained ~5 moles of the ligand per mole of the carrier, as determined by MALDI-TOF-MS. This discrepancy was attributed to the poor ionization of the glycosylated peptides. Analysis of the tryptic digests by MS/MS also showed the presence of numerous carbohydrate fragment ions which markedly complicated the MS/MS sequencing of the glycopeptides. However, it is recognized that the glycosylation of the carbohydrate haptens always occurs on lysine residues of the carrier BSA. Thus, we have previously proposed that the trypsin digestion of the glycoconjugate is not efficient to cleave the glycosylated lysines and this may affect the total digestion of the glycoconjugate.^[15]

In a separate study, we have investigated the location of glycation sites in an experimental neoglycoconjugate vaccine for anthrax by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-tandem mass spectrometry.^[16] The carbohydrate portion of the neoglycoconjugate was the synthetic tetrasaccharide side chain of the *Bacillus anthracis* exosporium, 2-*O*-methyl- β -D-glucopyranosyl-(1 \rightarrow 3)- α -L-rhamnopyranosyl-(1 \rightarrow 3)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranoside, and the protein carrier was BSA.^[16,17] The hapten:BSA ratio in the neoglycoconjugate was 5.4:1, as showed independently by MS. The conjugate was digested with either trypsin or GluC V8 before MS analysis. The advantage of using two different proteases results in more complete sequence coverage.

A total of 30 glycation sites were discovered when performing LC-ESI-QqTOF-MS/MS analysis of the digests, which is many more than was indicated by MALDI-TOF/TOF-CID-MS/MS analysis of the digests. Additionally, we also observed carbohydrate fragment ions characteristic of the tetrasaccharide portion.

Based on these observations, and to further explore power of MS in combination with enzymatic digestion in glycoconjugate analysis, here we use a glycoconjugate composed of a simple model carbohydrate, the milk sugar lactose, composed of two D-hexosyl residues and BSA as carrier. We hypothesize that the gas-phase reactivity of the β -lactose portion of the neoglycoconjugate should be different than that of the 4-(3-deoxy-L-*glycero*)-2-O-methyl- α -D-perosamine and of the tetrasaccharide 2-O-methyl- β -D-glucopyranosyl-(1 \rightarrow 3)- α -L-rhamnopyranosyl-(1 \rightarrow 3)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- α -L-rhamnopyranoside which we used previously.^[14,16] In addition, here we have used a combination of two different proteases, namely trypsin and the GluC V8, for the digestion of various glycoconjugates. It is well known that at pH 7.8, GluC V8 endoprotease cleaves peptide bonds at the C-terminus of glutamic acid and aspartic acid residues^[18,19]. Finally, the protease digests were analyzed using nano-LC-ESI-QqTOF-MS/MS.

6.2. Material and Methods

6.2.1. Hapten-BSA glycoconjugate preparation

Conjugation of β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranose (β -lactose) to BSA was carried out by the standard protocol.^[8] Briefly, commercial β -lactose octaacetate was converted to 5-methoxycarbonylpentyl β -lactoside which was transformed to a squaric acid monoester before conjugation. The conjugations were carried out in 0.5 M borate buffer (pH 9.0) at the initial hapten:BSA ratio of 20:1 and using BSA (Sigma Aldrich, Saint Louis, MO, USA) as carrier protein. The conjugation times were 3 and 24 hours.^[7] Isolation of the neoglycoconjugate was carried out by centrifugal ultrafiltration using filtration devices with a cut-off 30,000 Da.^[7]

It has been already established that when the foregoing protocol is applied, the carbohydrate-spacer moiety covalently attaches to the ϵ -amino group of the lysine residue in BSA.^[14,16] Figure 6.1 shows a representation of the synthesized carbohydrate hapten-BSA glycoconjugate.

6.2.2. Digestion

The digestions of the lactose-BSA glycoconjugates were carried out with trypsin and Glu-C V8 protease (Sigma Aldrich, Saint Louis, MO, USA). Thus, 100 μ g of the glycoconjugate was dissolved in a mixture of 0.1% RapiGest SF Surfactant (1 μ g, Waters, USA) in 50 mM of NH_4HCO_3 (100 μ L) at a pH of 8.0 and reduced by treatment with 2 μ L of 10 mM DTT (Sigma Aldrich, Saint Louis, MO, USA) for 30 min at room temperature, followed by alkylation with 2 μ L of a 50 mM iodoacetamide (Sigma Aldrich, Saint Louis, MO, USA) for 1 h at room temperature. A portion (50 μ g) of the

BSA glycoconjugate was digested with trypsin using a 20 ng/ml of trypsin dissolved in NH_4HCO_3 (50 mM, 1 mL) at a trypsin:glycoprotein ratio of 1:25 (w/w) and incubated at 37 °C overnight with shaking. The other 50 μg of the BSA glycoconjugate was digested using the GluC V8 endoprotease protease:glycoprotein ratio of 1:25 (w/w) and incubated at 37 °C overnight with shaking. The sample was then dried under vacuum and the residue was dissolved in 20 μL of 1% acetic acid (Sigma-Aldrich, Oakville, ON, Canada). An aliquot of each sample (10 μL) was then cleaned up using ZipTip C18 (Millipore, Bedford, MA, USA) before mass spectral analysis.

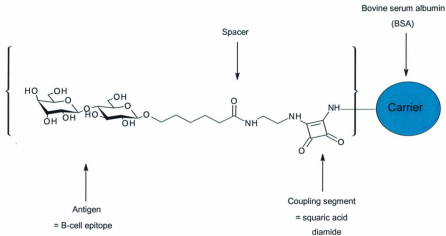


Figure 6.1. Schematic representation of the lactose-BSA glycoconjugate.

6.2.3. MALDI-TOF-MS analysis

Mass spectrometry analyses were carried out on an ABI 4800 with TOF-TOF optics (Applied Biosystems Foster City, CA) and a 200-Hz frequency Nd:YAG laser. α -Cyano-4-hydroxycinnamic acid was used as matrix for the analysis of BSA and hapten-BSA conjugates with an average of 5000 to 8000 laser shots per spectra. Briefly, 1 μ L of a 20 mg/mL solution of α -CHCA (dissolved in acetone, 0.1%TFA) was spotted on the MALDI plate and dried at room temperature. Then, an aliquot of 1 μ L of sample was spotted on the top of the dried matrix and let to dry before the MALDI-MS experiments. The analysis was achieved in the linear mode and the MALDI-TOF-MS was calibrated using BSA.

6.2.4. LC-ESI-QqTOF-MS/MS analysis

The peptides were separated on a DIONEX UltiMate3000 Nano LC System (Germering, Germany). 250 fmol of digested glycoprotein was dissolved in 0.1% TFA and loaded onto a precolumn (300 μ m i.d. \times 5 mm, C₁₈ PepMap100, 5 μ m (LC Packing, Sunnyvale, CA)) in order to desalt and concentrate the sample. After their elution from the precolumn, the peptides and glycopeptides mixtures were separated on a nanoflow analytical column (75 μ m i.d. \times 15 cm, C₁₈ PepMap 100, 3 μ m, 100 Å, (LC Packing, Sunnyvale, CA)) at a flow rate of 180 nL/min. The elution of the peptides and glycopeptides was achieved using the following mobile phases: 0.1% FA/0.01% TFA/2% ACN (A) and 0.08% FA/0.008% TFA/98% ACN (B). The elution started with 0% B for 10 min, followed by a gradient of 0-60% B in 55 min and 60-90% B in 3 min and was kept at 90% B for 3 min. The MS/MS analysis of the eluted peptides and glycopeptides was accomplished using an Applied Biosystems API-QSTAR XL quadrupole orthogonal

time-of-flight (QqTOF)-MS/MS hybrid tandem mass spectrometer (Applied Biosystems International-MDS Sciex, Foster City, CA, USA) equipped with a nanoelectrospray source (Protana XYZ manipulator) which produces the electrospray through a PicoTip needle (10 μm i.d., New Objectives, Woburn, MA, USA) carrying a voltage of 2400 V. The TOF analyzer was calibrated using a renin solution (1 pmol/ μL) and looking for the ions at m/z 586.9815 and m/z 879.9723. The collision energies used during the CID-MS/MS analyses were determined automatically using the Information Dependent Acquisition method integrated in the Analyst software.

6.3. Results

As previously indicated, the covalent attachment of lactose to BSA was carried out at an initial hapten:BSA ratio of 20:1 and the conjugation reaction was allowed to proceed for either 3 hours or 24 hours. Thus, two β -lactoside-BSA glycoconjugate vaccine models were produced.

6.3.1. MALDI-TOF-MS analysis of the hapten-BSA glycoconjugates.

In order to determine the average number of carbohydrate-spacer moieties linked to BSA (hapten:BSA ratio), the two glycoconjugates were analyzed using MALDI-TOF/TOF-MS (single stage). Molecular masses and hapten:BSA ratios were calculated by comparing the molecular mass of the conjugate to the molecular mass of the starting BSA. Thus, MALDI-TOF/TOF-MS revealed that the two neoglycoconjugates formed had hapten:BSA ratios 5.1:1 (reaction time: 3 hours) and 19.0:1 (reaction time: 24 hours). These hapten:BSA ratios were found to be identical to those previously determined by SELDI-TOF-MS.^[7]

The molecular mass calculations of these glycoconjugates were based on the presumption that squaric acid covalently links to the ϵ -amino groups of the lysine residues in BSA by loss of an ethanol molecule^[13,15] and that the molecular mass of the fragment of disaccharide-spacer-squaric acid which will bond to the BSA is 577.22 Da (Table 6.1).

Figure D.1 (Appendix D) and Table 6.1 display the mass spectra and the m/z values of (a) BSA, (b) the synthetic neoglycoconjugate vaccine obtained after a reaction time of 3 hours and (c) the synthetic neoglycoconjugate vaccine obtained after a reaction time of 24 hours.

Table 6.1. Molecular ions identified during the MALDI-TOF/TOF-MS analysis of the lactose-BSA glycoconjugates obtained with different reaction times.

Reaction time	Observed ion m/z		Calculated Lactose:BSA ratio
	$[M + 2H]^{2+}$	$[M + H]^+$	
BSA	33161.81	66318.88	-
3 h	34619.44	69307.96	5.1
24 h	38683.93	77258.62	19.0

6.3.2. LC-ESI-QqTOF-MS/MS analyses of the tryptic and GluC V8 digests of the neoglycoconjugate with a lactose:BSA ratio of 5.1:1 and 19.0:1

Nano-LC-ESI-QqTOF-MS/MS analysis was carried out on both tryptic and GluC V8 digests of glycoconjugates with a hapten:protein ratio of 5.1:1 and 19.0:1. The MS/MS spectra obtained were submitted to the Mascot library in order to identify the peptides formed from BSA. The Mascot reports of LC-ESI-QqTOF-MS/MS data of the tryptic digest of the glycoconjugates demonstrated the identification of two serum albumin isoforms from the *Bos taurus* species. The serum albumin precursor ([gi|1351907](#)) was identified with the following sequence coverage: 45% and 47%, for the neoglycoconjugates with a lactose:BSA ratio of respectively 5.1:1 and 19.0:1. The serum albumin protein ([gi|74267962](#)) was also identified for the analysis of the tryptic digests of the glycoconjugates with a hapten:protein ratio of 5.1:1 (sequence coverage = 41%) and 19.0:1 (sequence coverage = 43%).

Similarly, the data of the LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests of the glycoconjugates with ratios of 5.1:1 and 19.0:1 were submitted to the Mascot library which gave a match to the same two serum albumin proteins from the *Bos taurus* species mentioned previously: the serum albumin precursor ([gi|1351907](#)) and the serum albumin ([gi|74267962](#)). The sequence coverage for the albumin precursor ([gi|1351907](#)) for the neoglycoconjugates with hapten:BSA ratios of 5.1:1 and 19.0:1 was found to be 39% and 37%, respectively. Finally, serum albumin protein ([gi|74267962](#)) was identified with the following sequence coverage: 40% for the hapten-BSA glycoconjugate 5.1:1 and 33% for the hapten-BSA glycoconjugate 19.0:1.

Most importantly, the reports obtained also contained a list of ions that did not match any protein; these were considered to be potentially glycosylated BSA peptides. These ions were selected and subjected to CID-MS/MS sequencing, in order to identify the correct sequences of the glycosylated peptides and their exact glycosylation sites.

6.3.3. Mass spectrometry determination of the glycosylation sites on the neoglycoconjugate with a hapten:BSA ratio of 5.1:1

LC-ESI-QqTOF-MS/MS analysis of the tryptic digests

The LC-ESI-QqTOF-MS/MS analysis of the tryptic digests of the neoglycoconjugate with a hapten:BSA ratio of 5.1:1 allowed us to identify nine glycopeptides having lysine residues occupied on different glycosylation sites designed with an asterisk. These were identified as follows (Table 6.2): SLGK*VGTR (Lys 455) at m/z 697.3748 (+2), LSQK*FPK (Lys 245) at m/z 712.3812 (+2), EK*VLTSSAR (Lys 211) at m/z 783.8877 (+2), ALK*AWSVAR (Lys 235) at m/z 789.4228 (+2), LAK*EYEATLEECCA (Lys 374) at m/z 797.7141 (+3), CASIQK*FGER (Lys 228) at m/z 886.4299 (+2), CCTK*PESER (Lys 463) at m/z 871.8738 (+2), VTK*CCATESLVNR (Lys 498) at m/z 1021.9941 (+2) and K*VPQVSTPTLVEVSR (Lys 437) at m/z 1108.5900 (+2). Figure 6.2 displays the following examples of different glycopeptides obtained by tryptic digests of the lactose-BSA glycoconjugate: (a) SLGK*VGTR (Lys 455) at m/z 697.3748 (+2) and (b) LSQK*FPK (Lys 245) at m/z 712.3812 (+2), (c) EK*VLTSSAR (Lys 211) at m/z 783.8877 (+2) and (d) ALK*AWSVAR (Lys 235) at m/z 789.4228 (+2). The CID-MS/MS fragmentation pathways obtained by the isolated precursor glycopeptide ions enabled us to unambiguously determine the various glycosylation sites.

Table 6.2. Tryptic glycopeptides identified in the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the lactose-BSA glycoconjugates.

Peptide		Missed cleavage	Lactose:BSA 5.1:1		Lactose:BSA 19.0:1	
Sequence (asterisk = glycation site)	Calculated <i>m/z</i> (charge)		Observed <i>m/z</i> (charge)	Deviation (Da)	Observed <i>m/z</i> (charge)	Deviation (Da)
SLGK*VGTR (Lys 455)	697.3568 (+2)	1	697.3748 (+2)	0.0181	697.3567 (+2)	-0.0001
LSQK*FPK (Lys 245)	712.3640 (+2)	2	712.3812 (+2)	0.0172	712.3673 (+2)	0.0033
EK*VLTSSAR (Lys 211)	783.8912 (+2)	1	783.8877 (+2)	-0.0035	783.8900 (+2)	-0.0012
ALK*AWSVAR (Lys 235)	789.4068 (+2)	1	789.4228 (+2)	0.0161	789.4125 (+2)	0.0058
LAK*EYEATLEECCA (Lys 374)	797.6874 (+3)	2	797.7141 (+3)	0.0267	797.6954 (+3)	0.0080
CASIQK*FGER (Lys 228)	886.4066 (+2)	1	886.4299 (+2)	0.0233	886.4033 (+2)	-0.0033
CCTK*PESER (Lys 463)	871.8587 (+2)	1	871.8738 (+2)	0.0151	871.8000 (+2)	-0.0587
VTK*CCTESLVNR (Lys 498)	1021.9667 (+2)	1	1021.9941 (+2)	0.0274	1021.9741 (+2)	0.0074
K*FWGK (Lys 156)	621.3007 (+2)	1	-	-	-	-
TPVSEK*VTK (Lys 495)	782.8959 (+2)	2	-	-	782.8995 (+2)	0.0036
K*VPQVSTPTLVEVSR (Lys 437)	1108.5811 (+2)	0	-	-	1108.5854 (+2)	0.0043
SLHTLFGDELCK*VASLR (Lys 100)	841.4162 (+3)	1	-	-	841.4000 (+3)	-0.0162

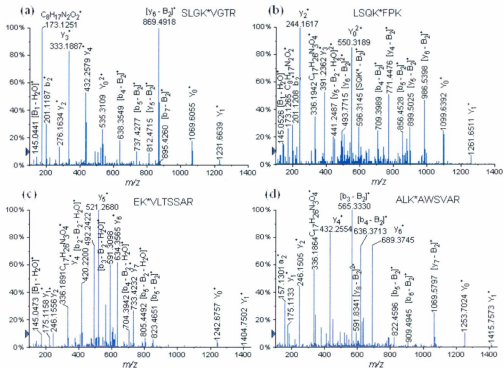


Figure 6.2. LC-QqTOF-MS/MS spectra of the tryptic glycosylated peptides (a) SLGK*VGTR (Lys 455) at m/z 697.3567 (+2), (b) LSQK*FPK (Lys 245) at m/z 712.3673 (+2), (c) EK*VLTSSAR (Lys 211) at m/z 783.8900 (+2) and (d) ALK*AWSVAR (Lys 235) at m/z 789.4125 (+2).

During the CID-MS/MS analyses, we used the pre-existing peptide nomenclature involving the detection of the x-, y-, z-ions and the a-, b-, c-ions for the individual peptide sequence.^[20,21] For the glycopeptide product ions, we used the nomenclature established by Domon and Costello for the carbohydrate portion as: A, B, C, X, Y and Z.^[22]

The CID-MS/MS analysis of the doubly charged glycated peptide SLGK*VGTR (Lys 455) at m/z 697.3748 (Figure 6.2(a), Table D.1) afforded a series of product ions corresponding to the entire peptide chain formed by loss of the β -D-galactopyranosyl moiety: Y_1^+ at m/z 1231.6639, and the entire β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside: Y_0^+ at m/z 1069.6055, Y_0^{2+} at m/z 535.3109 and $[Y_0 - H_2O]^{2+}$ at m/z 526.3004 (representation in Figure 6.3). Additionally, we have noted the formation of product ions corresponding to the galactopyranosyl moieties: B_1^+ at m/z 163.0556, $[B_1 - H_2O]^+$ at m/z 145.0441 and $[B_1 - 2H_2O]^+$ at m/z 127.0377. Clearly, these carbohydrate fragments confirmed the identity of the carbohydrate conjugated to the BSA and were an epitome of chemical stability of these cations in gas phase. We also noted the formation of the product ions corresponding to the squaric acid-spacer: $C_{17}H_{26}N_3O_4^+$ at m/z 336.1908, $C_{11}H_{16}N_3O_2^+$ at m/z 222.1271 and $C_8H_{17}N_2O_2^+$ at m/z 173.1251 (product ions displayed in Figure 6.3). As expected, we observed a propensity of the gas-phase loss of the carbohydrate disaccharide portion during the MS/MS analysis. For example, the following peptide product ions were observed and formed by the loss of the carbohydrate moiety: $[b_7 - B_2]^+$ at m/z 895.4260, $[y_6 - B_2]^+$ at m/z 869.4918, $[y_5 - B_2]^+$ at m/z 812.4715, $[b_6 - B_2]^+$ at m/z 794.4752, $[b_5 - B_2]^+$ at m/z 737.4277, $[b_5 - B_2 - H_2O]^+$ at m/z 719.3841, $[a_5 - B_2]^+$ at m/z 709.4391, $[b_4 - B_2]^+$ at m/z 638.3549 and $[y_6 - B_2 - H_2O]^{2+}$ at m/z 426.2460.

Finally, we also noted the presence of the conventional y- and b- product ions corresponding to the fragmentation of the peptide backbone, which allowed us to cover the sequence of each entire glycated peptide.

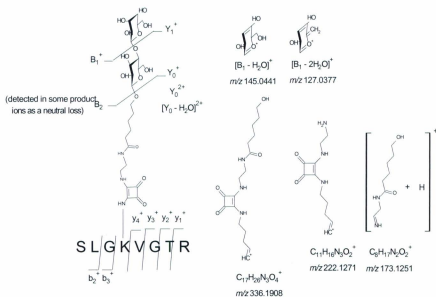


Figure 6.3. Product ions resulting from the tandem mass spectrometry analysis of the glycosylated peptide SLGK*VGTR (Lys 455) at m/z 697.3567 (+2) extracted from the LC-ESI-Qq-TOF-MS/MS analysis of the tryptic digests of the hapten-BSA glycoconjugate with a lactose:BSA ratio of 5.1:1.

LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests

The manual sequencing of the spectra that were not assigned to a peptide of the *B. taurus* protein on the Mascot report, enabled us to identify eight glycopeptides containing different lysine residue glycation sites at different lysine residues (Table 6.3): K*TPVSE (Lys 489) at m/z 618.7989, HVK*LVNE (Lys 65) at m/z 707.8591, K*LFTFHAD (Lys 528) at m/z 777.9566 (+2), LCK*VASLRE (Lys 100) at m/z 826.4216 (+2), VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7418 (+3), VSRSLGK*VGTRCCTKPESE (Lys 455) at m/z 909.7741 (+3), DKGACLLPK*IE (Lys 204) at m/z 910.4582 (+2), YAVSVLLRLAK*E (Lys 374) at m/z 969.5187 (+2), LLYYANK*YNGVFQE (Lys 183) at m/z 1149.5413 (+2) and K*LFTFHADICTLPDTE (Lys 528) at m/z 1242.5946 (+2) (Table 6.3). Figure 6.4(a) and 6.4(b) display several examples of the identified glycopeptides during the LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests: (a) VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7418 (+3) and (b) DKGACLLPK*IE (Lys 204) at m/z 910.4582 (+2).

The MS/MS analysis of the precursor glycopeptide ion VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7418 (+3) shown in Figure 6.4(a) (Table D.2) adopted a CID-fragmentation pathway similar to the other tryptic digested glycopeptides ions. These CID-MS/MS fragmentations appear to be initiated by loss of the entire disaccharide moiety. Thus the following product ions were identified: the entire peptide ion produced by loss of the β -D-galactopyranosyl moiety, the product ion Y_1^{2+} at m/z 1175.0537 and the intact disaccharide product (β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside) ion, Y_0^{2+} at m/z 1094.0196 and Y_0^{3+} at m/z 729.6955. It is important to point out the formation of the diagnostic product ions, which were created specifically by

the carbohydrate-spacer moiety: the mixed product ions containing a fragment of lysine and the squaric acid-spacer assigned as $[C_{17}H_{26}N_3O_4]^+$ at m/z 336.1871 and $[C_{11}H_{16}N_3O_2]^+$ at m/z 222.1181. In addition, we were also able to detect the product ions corresponding to monosaccharide galactopyranosyl moiety: B_1^+ at m/z 163.0527 and $[B_1 - H_2O]^+$ at m/z 145.0423.

Additionally, some peptide fragment ions formed by loss of the entire β -lactose moiety were also detected: $[b_{13} - B_2]^{2+}$ at m/z 857.3536, $[b_{16} - B_2]^{3+}$ at m/z 680.9973 and $[b_{15} - B_2]^{3+}$ at m/z 648.3185. Furthermore, other product ions covering the sequence of the glycopeptide are displayed in Figure 6.4(a) and Table D.2. Another example of a glycopeptide identified during the LC-ESI-QqTOF-MS/MS analysis of GluC V8 digests of the precursor DKGACLLPK*IE ion (Lys 204) at m/z 910.4582 (+2) is shown in Figure 6.4(b).

Table 6.3. Glycopeptides identified in the bovine serum albumin protein by LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests of the lactose-BSA glycoconjugates.

Peptide		Missed cleavage	Lactose:BSA 5.1:1		Lactose:BSA 19.0:1	
Sequence (star = glycation site)	Calculated <i>m/z</i> (charge)		Observed <i>m/z</i> (charge)	Deviation (Da)	Observed <i>m/z</i> (charge)	Deviation (Da)
K*TPVSE (Lys 489)	618.7904 (+2)	0	618.7989 (+2)	0.0085	618.7940 (+2)	0.0036
HVK*LVNE (Lys 65)	707.8513 (+2)	0	707.8591 (+2)	0.0078	707.8565 (+2)	0.0052
K*LFTFHAD (Lys 528)	777.8644 (+2)	0	777.9566 (+2)	0.0922	777.9564 (+2)	0.092
LCK*VASLRE (Lys 100)	826.4087 (+2)	0	826.4216 (+2)	0.0129	826.4137 (+2)	0.005
VSRSLGK*VGTRCCTKPE (Lys 455)	837.7407 (+3)	0	837.7418 (+3)	0.0011	837.7468 (+3)	0.0061
VSRSLGK*VGTRCCTKPESE (Lys 455)	909.7656 (+3)	1	909.7741 (+3)	0.0085	909.7714 (+3)	0.0058
DKGACLLPK*IE (Lys 204)	910.4480 (+2)	1	910.4582 (+2)	0.0103	910.4539 (+2)	0.0059
YAVSVLLRLAK*E (Lys 374)	969.5198 (+2)	0	969.5187 (+2)	-0.0011	969.4790 (+2)	-0.0408
LLYYANK*YNGVFQE (Lys 183)	1149.5389 (+2)	0	1149.5413 (+2)	0.0024	1149.5565 (+2)	0.0176
K*LFTFHADICTLPDTE (Lys 528)	1242.5726 (+2)	2	1242.5946 (+2)	0.022	1242.5876 (+2)	0.0150
K*SHCIAE (Lys 309)	710.8113 (+2)	0	-	-	710.8039 (+2)	-0.0074
K*SLHTLFGD (Lys 88)	797.3804 (+2)	0	-	-	797.3361 (+2)	-0.0443
CCDK*PLLE (Lys 304)	805.8445 (+2)	1	-	-	805.8449 (+2)	0.0004
FAK*TCVADE (Lys 75)	808.8481 (+2)	1	-	-	808.8551 (+2)	0.007
GPK*LVVSTQTALA (Lys 597)	930.9883 (+2)	0	-	-	931.0042 (+2)	0.0159
IAHRFK*DLGEE (Lys 36)	945.9523 (+2)	2	-	-	945.9632 (+2)	0.0109
AK*DAFLGSFLYE (Lys 346)	968.9514 (+2)	1	-	-	968.9538 (+2)	0.0024
VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463)	1029.8153 (+3)	0	-	-	1029.8239 (+3)	0.0086
K*DAIPE (Lys 318)	624.7904 (+2)	1	-	-	624.8000 (+2)	0.0096
VTK*LVTD (Lys 256)	676.3402 (+2)	0	-	-	676.3457 (+2)	0.0055

K*SHCIAEVE (Lys 309)	824.8668 (+2)	1	-	-	824.8706 (+2)	0.0038
LTK*VHK*E (Lys 263, Lys 266)	1003.9750 (+2)	0	-	-	1003.9831 (+2)	0.0081
K*VTK*CCTE (Lys 495, Lys 498)	1089.4619 (+2)	0	-	-	1089.4600 (+2)	-0.0019
VSRSLGK*VGTRCCTK*PESE (Lys 455, Lys 463)	1101.8402 (+3)	1	-	-	1101.8476 (+3)	0.0074
DK*GACLLPK*IE (Lys 197, Lys 204)	1198.5599 (+2)	1	-	-	1198.5604 (+2)	0.0005
KK*FWGK*YLYE (Lys 156, 160)	1257.5868 (+2)	0	-	-	1257.5983 (+2)	0.0115

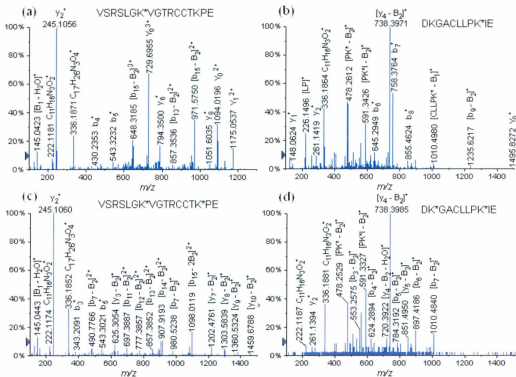


Figure 6.4. LC-QqTOF-MS/MS spectra of the GluC V8 digests glycosylated peptides (a) VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7418 (+3), (b) DKGACLLPK*IE (Lys 204) at m/z 910.4582 (+2), (c) VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 1029.8189 (+3) and (d) DK*GACLLPK*IE (Lys 197, Lys 204) at m/z 1198.5604 (+2).

Tus, a total of 15 glycation sites were identified during the LC-ESI-QqTOF-MS/MS analysis of tryptic and GluC V8 digests of the lactose-BSA glycoconjugate, which are displayed in Figure 6.5(a). When we compare the average number of glycation sites determined by the foregoing analysis (15 glycation sites) to the ratio of 5.1:1 determined for the lactose-BSA conjugate, the difference can be explained by the formation of a mixture of glycoforms during conjugation. Finally, the total sequence coverage of the glycoconjugate obtained by the trypsin and GluC V8 digestion was found to be 86%.

(a)

```

1 MKWVTFISLL LLFSSAYSRG VFRDTHKSE IAHRFKDLGE EHFKGLVLIA
51 FSQYLQCCPF DEHVK*LVNEL TEFAK*TCVAD ESHAGCEKSL HTLFGDELCK*
101 VASIRETYGD MADCCAKQEP ERNECFLSHK DSDPDLPKLK PDNTLCDEF
151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANK*YNGVFQE CQQAEDK*GAC
201 LLPK*IETMRE K*VLTSSARQR LRCASIQK*FG ERALK*AWSVA RLSQK*FPKAE
251 FVEVTKLVTD LTKVHKECCH GDLECCADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAK*EYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCQDFEK LGEYGFQNEL IVRYTRK*VPQ VSTPTLVEVS
451 RSLGK*VGTRC CTK*PESERMP CAEDYLSLIL NRLCVLHEK*T PVSEKVT*CC
501 TESLVNRRPC FSALTPDETY VPKAFDEK*LF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVGK CCAADDKEAC FAVEGPKLVV
601 STQTALA

```

(b)

```

1 MKWVTFISLL LLFSSAYSRG VFRDTHKSE IAHRFK*DLGE EHFKGLVLIA
51 FSQYLQCCPF DEHVK*LVNEL TEFAK*TCVAD ESHAGCEK*SL HTLFGDELCK*
101 VASIRETYGD MADCCAKQEP ERNECFLSHK DSDPDLPKLK PDNTLCDEF
151 KADEKK*FWGK* YLYEIARRHP YFYAPELLYY ANK*YNGVFQE CQQAEDK*GAC
201 LLPK*IETMRE K*VLTSSARQR LRCASIQK*FG ERALK*AWSVA RLSQK*FPKAE
251 FVEVTK*LVTD LTK*VHK*ECCH GDLECCADDR ADLAKYICDN QDTISSKLKE
301 CCDK*PLLEK*S HCIAEVEK*DA IPENLPPLTA DFAEDKDVCK NYQEAK*DAFL
351 GSFLYEYSRR HPEYAVSVLL RLAK*EYEATL EECCAKDDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCQDFEK LGEYGFQNEL IVRYTRK*VPQ VSTPTLVEVS
451 RSLGK*VGTRC CTK*PESERMP CAEDYLSLIL NRLCVLHEK*T PVSEK*VTK*CC
501 TESLVNRRPC FSALTPDETY VPKAFDEK*LF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVGK CCAADDKEAC FAVEGPK*LVV
601 STQTALA

```

Figure 6.5. BSA sequence where the glycation sites are indicated by an asterisk (red = identified on tryptic digests, blue = identified on GluC V8 digests and red and underlined = identified on both tryptic and GluC V8 digests) for the neoglycoconjugates with a lactose:BSA ratio of (a) 5.1:1 and (b) 19:0.1.

6.3.4. Mass spectrometry determination of the glycation sites on the neoglycoconjugate with a hapten:BSA ratio of 19.0:1

LC-ESI-QqTOF-MS/MS analysis of the tryptic digests

The following glycated peptides were identified allowing the determination of 11 glycation sites (Table 6.2) and these were assigned as follows: SLGK*VGTR (Lys 455) at m/z 697.3567 (+2), LSQK*FPK (Lys 245) at m/z 712.3673 (+2), TPVSEK*VTK (Lys 495) at m/z 782.8995 (+2), EK*VLTSSAR (Lys 211) at m/z 783.8900 (+2), ALK*AWSVAR (Lys 235) at m/z 789.4125 (+2), LAK*EYEATLEECCA (Lys 374) at m/z 797.6954 (+3), SLHTLFGDELCK*VASLR (Lys 100) at m/z 841.4000 (+3), CASIQK*FGER (Lys 228) at m/z 886.4033 (+2), CCTK*PESER (Lys 463) at m/z 871.8000 (+2), VTK*CTESLVNR (Lys 498) at m/z 1021.9741 (+2) and K*VPQVSTPTLVEVSR (Lys 437) at m/z 1108.5854 (+2).

LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests

The LC-ESI-QqTOF-MS/MS analysis of the GluC V8 digests of the neoglycoconjugate with a hapten:BSA ratio of 19.0:1 permitted us to determine 24 glycation sites through the identification of the following glycated peptides (Table 6.3): K*TPVSE (Lys 489) at m/z 618.7940 (+2), K*DAIPE (Lys 318) at m/z 624.8000 (+2), VTK*LVTD (Lys 256) at m/z 676.3457 (+2), HVK*LVNE (Lys 65) at m/z 707.8565 (+2), K*SHCIAE (Lys 309) at m/z 710.8039 (+2), K*LFTFHAD (Lys 528) at m/z 777.9564 (+2), K*SLHTLFGD (Lys 88) at m/z 797.3361 (+2), CCDK*PLLE (Lys 304) at m/z 805.8449 (+2), FAK*TCVADE (Lys 75) at m/z 808.8551 (+2), K*SHCIAEVE (Lys 309) at m/z 824.8706 (+2), LCK*VASLRE (Lys 100) at m/z 826.4137 (+2), VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7468 (+3),

VSRSLGK*VGTRCCTKPESE (Lys 455) at m/z 909.7714 (+3), DKGACLLPK*IE (Lys 204) at m/z 910.4539 (+2), GPK*LVVSTQTALA (Lys 597) at m/z 931.0042 (+2), IAHRFK*DLGEE (Lys 36) at m/z 945.9632 (+2), AK*DAFLGSFLYE (Lys 346) at m/z 968.9538 (+2), YAVSVLLRLAK*E (Lys 374) at m/z 969.4790 (+2), LLYYANK*YNGVFQE (Lys 183) at m/z 1149.5565 (+2) and K*LFTFHADICTLPDTE (Lys 528) at m/z 1242.5876 (+2).

In this case, it is worth mentioning that we were able to scrutinize 6 glycopeptides which carry two nearby glycation sites. These were assigned as follows: LTK*VHK*E (Lys 263, Lys 266) at m/z 1003.9831 (+2), VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 1029.8239 (+3), K*VTK*CCTE (Lys 495, Lys 498) at m/z 1089.4600 (+2), VSRSLGK*VGTRCCTK*PESE (Lys 455, Lys 463) at m/z 1101.8476 (+3), DK*GACLLPK*IE (Lys 197, Lys 204) at m/z 1198.5604 (+2) and KK*FWGK*YLYE (Lys 156, 160) at m/z 1257.5983 (+2). An interesting example to discuss is the diagnostic peptide VSRSLGKVGTRCCTKPE, which was detected with either one single glycation site on the Lys 455 residue affording the glycopeptide at m/z 837.7468 (+3) or with two glycation sites on the Lys 455 and Lys 463 residues, producing the glycosylated peptide at m/z 1029.8239 (+3). The CID-MS/MS analysis of the precursor ions containing two glycation sites VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 1029.8239 (+3) is shown in Figure 6.4 (c) and Table D.3. The CID-fragmentation pathway of this precursor glycosylated peptide ion afforded peptide product ions formed by loss of only one entire β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside. Thus, we detected the y-product ions obtained by the loss of the entire carbohydrate moiety on the Lys 463 residue. The majority of the remaining product ions were identified as follows: $[y_{10} - B_2]^+$

at m/z 1459.6788, $[y_9 - B_2]^+$ at m/z 1360.5324, $[y_8 - B_2]^+$ at m/z 1303.5839, $[y_7 - B_2]^+$ at m/z 1202.4761, $[y_4 - B_2]^+$ at m/z 726.3182 and $[y_3 - B_2]^+$ at m/z 625.3054. However, it is interesting to note that the loss of the entire β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside on the Lys 455 residue was also observed on the following b-product ions: $[b_8 - B_2]^+$ at m/z 1079.6271, $[b_7 - B_2]^+$ at m/z 980.5238, $[b_{14} - B_2]^{2+}$ at m/z 907.9193, $[b_{13} - B_2]^{2+}$ at m/z 857.3852, $[b_{12} - B_2]^{2+}$ at m/z 777.3857, $[b_{11} - B_2]^{2+}$ at m/z 697.3887, $[b_9 - B_2]^{2+}$ at m/z 568.7979 and $[b_7 - B_2]^{2+}$ at m/z 490.7766. Finally, during the product ion scan of the precursor VSRLGK*VGTRCCTK*PE ion (Lys 455, Lys 463) at m/z 1029.8239 (+3), we identified a product ion formed by the loss of the two β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranosides carried on Lys 455 and Lys 463 residues: $[b_{15} - 2B_2]^{2+}$ at m/z 1098.0119. In addition, another series of diagnostic product ions, which were described earlier were detected and assigned as: $[C_{17}H_{26}N_3O_4]^+$ at m/z 336.1852 and $[C_{11}H_{16}N_3O_2]^+$ at m/z 222.1174. The galactopyranosyl $[B_1]^+$ product ion at m/z 163.0559 was also observed.

Thus, it is important to note, that the observation of peptides having the same sequence which were glycosylated with either one and/or two carbohydrate haptens, confirms the fact that the formed glycoconjugate was indeed a mixture of glycoforms.

Figure 6.4(d) displays the LC-ESI-QqTOF-MS/MS analysis of glycopeptide DK*GACLLPK*IE at m/z 1198.5604 (+2), derivatized at the Lys 197 and Lys 204 residues, and the product ions with their m/z values are reported in Table D.4. Moreover, the peptides with the following sequences VSRLGKVGTRCCTKPE and VSRLGKVGTRCCTKPESE were also observed with one glycosylation on the Lys 455 residue and two glycosylations on the Lys 455 and Lys 463 residues (Table 6.3).

When combining the data of the LC-ESI-QqTOF-MS/MS analysis of the tryptic and GluC V8 digests, a total of 30 glycation sites were identified [displayed in Figure 6.5(c)], and the total sequence coverage was evaluated to be 89 %.

Figure 6.6 displays the 3-D structure (obtained using Swiss-Prot pdb viewer^[23,24]) of the lactose-BSA glycoconjugates with a hapten:BSA ratio of (a) 5.1:1 and (b) 19.0:1, where the glycated lysines are highlighted in red. We can thus observe that the majority of the glycated lysine residues are located on the outer-surface of the protein.

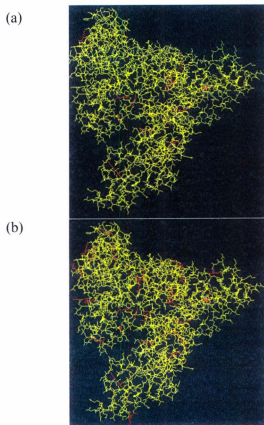


Figure 6.6. 3-D structure of the lactose-BSA glycoconjugates with a lactose:BSA ratio of (a) 5.1:1 and (b) 19.0:1.

6.4 Conclusion

The MALDI-TOF/TOF-MS analysis of the different lactose-BSA glycoconjugates allowed us to determine the different hapten:BSA ratios formed during conjugation: 5.1:1 and 19.0:1. The LC-ESI-QqTOF-MS/MS analysis identified 15 glycation sites for the neoglycoconjugate with a hapten:BSA ratio of 5.1:1 and 30 glycation sites for the neoglycoconjugate with a hapten:BSA ratio of 19.0:1, showing that the conjugates were mixtures of glycoforms. This could be deduced from a higher number of glycation sites compared to the average number of haptens present in conjugates per mol of the carrier. The presence of glycoforms was also confirmed by the fact that for the same glycoconjugate with a hapten:BSA ratio of 19.0:1, glycopeptides of the same amino acid sequence were detected containing either one or two different glycation sites.

As expected, it was observed that the number of identified glycation sites increased when the hapten:BSA ratio of glycoconjugate formation increased and that the location of the glycation sites appears to be mainly on the outer-surface of the BSA carrier molecule (Figure 6.6), which is in line with the assumption that the sterically more accessible lysine residues, namely those located on the outer-surface of the BSA, would be conjugated preferentially.

As also observed during the LC-ESI-QqTOF-MS/MS analysis of the digested *Bacillus anthracis* exosporium specific tetrasaccharide-BSA glycoconjugate,^[16] we were able to detect various conventional diagnostic carbohydrate fragment ions ($[B_1]^+$, $[B_1 - H_2O]^+$ and $[B_1 - 2H_2O]^+$). We also noted the formation of Y_0^+ product ions and, in some glycopeptides, the loss of the neutral B_2 product ions (formation of $[y - B_2]^+$ and $[b - B_2]^+$

ions). These findings seem to indicate the gas-phase fragility of these ions and inclination of the glycosylated oxygen-carbon spacer bond to rupture. Therefore, we conclude that regardless of the nature the carbohydrate portion in the synthetic neoglycoconjugates under investigation, the glycosylated O-6-C-1 bond, between the aglycone and the carbohydrate portion, is easy to rupture in the gas-phase.

Appendix D. Supplementary data

Supplementary data are associated with this chapter (page 267)

References

- [1] P. Kováč. Synthetic oligosaccharides. Indispensable probes in the life sciences, Ed.; ACS Symposium Series 560; American Chemical Society: Washington, DC, **1994**.
- [2] Y. C. Lee, R. T. Lee. In *Synthetic glycoconjugates*; H. J. Allen, E. C. Kisailus, Eds.; Marcel Dekker: New York, **1992**, p. 121.
- [3] D. P. Galonic, D. Y. Gin. Chemical glycosylation in the synthesis of glycoconjugate antitumor vaccines. *Nature* **2007**, *446*, 1000.
- [4] W. E. Dick Jr., M. Beurret. Glycoconjugates of Bacterial Carbohydrate Antigens. In *Conjugate vaccines*. J. M. Cruise, R. E. Jr. Lewis (Eds.) Switzerland: Basel, Krager, vol. 10, **1989**, p. 48.
- [5] L. F. Tietze, M. Arlt, M. Beller, K. H. Glusenkamp, E. Jahde, M. F. Rajewsky. Anticancer agents, 15. Squaric acid diethyl ester: a new coupling reagent for the formation of drug biopolymer conjugates. Synthesis of squaric acid ester amides and diamides. *Chem. Ber.* **1991**, *124*, 1215.
- [6] V. P. Kamath, P. Diedrich, O. Hindsgaul. Use of diethyl squarate for the coupling of oligosaccharide amines to carrier proteins and characterization of the resulting neoglycoproteins by MALDI-TOF mass spectrometry. *Glycoconj. J.* **1996**, *13*, 315.
- [7] S. J. Hou, R. Saksena, P. Kováč. Preparation of glycoconjugates by dialkyl squarate chemistry revisited. *Carbohydr. Res.* **2008**, *343*, 196.
- [8] J. Zhang, A. Yergey, J. Kowalak, P. Kováč. Studies towards neoglycoconjugates from the monosaccharide determinat of *Vibrio cholerae* O:1 serotype Ogawa, using the diethyl squarate reagent. *Carbohydr. Res.* **1998**, *313*, 15.

- [9] R. Saksena, X. Ma, P. Kováč. One-pot preparation of a series of glycoconjugates with predetermined antigen-carrier ratio from oligosaccharides that mimic the O-PS of *Vibrio cholerae* O:1, serotype Ogawa. *Carbohydr. Res.* **2003**, 338, 2591.
- [10] A. Chernyak, A. Karavanov, Y. Ogawa, P. Kováč. Conjugating oligosaccharides to proteins by squaric acid diester chemistry: rapid monitoring of the progress of conjugation, and recovery of the unused ligand. *Carbohydr. Res.* **2001**, 330, 479.
- [11] X. Ma, R. Saksena, A. Chernyak, P. Kováč. Neoglycoconjugates from synthetic tetra- and hexasaccharides that mimic the terminus of the O-PS of *Vibrio cholerae* O:1, serotype Inaba. *Org. Biomol. Chem.* **2003**, 1, 775.
- [12] R. Saksena, A. Chernyak, A. Karavanov, P. Kováč, in *Methods Enzymol.* Vol. 362, Y. C. Lee and R. Lee (eds), Academic press, New York, **2003**, p. 125.
- [13] Hegedűs G, Bélai I, Székács A. Development of an enzyme-linked immunosorbent assay (ELISA) for the herbicide trifluralin. *Anal. Chim. Acta* **2000**, 421, 121.
- [14] F. Jahouh, R. Saksena, D. Aiello, A. Napoli, G. Sindona, P. Kováč, J. H. Banoub. Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry. *J. Mass Spectrom.* **2010**, 45, 1148.
- [15] N. D. Rawlings, A. J. Barrett. Families of serine peptidases. *Methods Enzymol.* **1994**, 244, 19.
- [16] F. Jahouh, S. J. Hou, P. Kováč, Banoub J. H. Determination of the glycation sites of *Bacillus anthracis* neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-tandem mass spectrometry. *J. Mass Spectrom.* **2011**, 46, 993.

- [17] J. M. Daubenspeck, H. Zeng, P. Chen, S. Dong, C. T. Steichen, N. R. Krishna, D. G. Pritchard, C. L. Turnbough. *J. Biol. Chem.* **2004**, 279, 30945.
- [18] G. R. Drapeau, Y. Boily, J. Houmard. Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J. Biol. Chem.* **1972**, 247, 6720.
- [19] J. J. Birktoft, K. Breddam. Proteolytic enzymes: Glutamyl endopeptidases. *Methods Enzymol.* **1994**, 244, 114.
- [20] P. Roepstorff, J. Fohlman. Letter to the editors. *Biol. Mass Spectrom.* **1984**, 11, 601.
- [21] R.S. Johnson, S. A. Martin, K. Biemann, J. T. Stults, J. T. Watson. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, 59, 2621.
- [22] B. Domon, C. Costello. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, 5, 397.
- [23] K. Arnold, L. Bordoli, J. Kopp, T. Schwede. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, 22, 195.
- [24] M. C. Peitsch. Protein modeling by E-mail. *Biol. Technol.* **1995**, 13, 658.

CHAPTER 7: General conclusion

In this study, tandem mass spectrometry has been used for the analysis of oligonucleotides and synthetic carbohydrate-protein neoglycoconjugate vaccines.

The ESI-QqTOF-MS analysis of the three constitutional isobaric 18-mer DNA oligomers (GATTCATAGCTACGAATC **1**, AATTCGTAGCTACGAATC **2**, and AATTCGTACCTACGAATG **3**) produced the same multiply-charged molecular ions. We have shown that the CID-MS/MS analysis of the deprotonated molecular ions $[M - 8H]^{8-}$ at m/z 683.9881 and $[M - 9H]^{9-}$ at m/z 607.8775 permitted us to establish their sequences and the verification of the product ions obtained were confirmed by using the Mongo Oligo Mass Calculator. Furthermore, we reported that the CID-MS/MS of the precursor ions isolated from each distinct oligonucleotide created characteristic product ions which permitted their differentiation and characterization. The CID-MS/MS studies allowed us to define the mass spectrometry fingerprint of these three isobaric oligonucleotides. This described work on these three oligonucleotides, was also undertaken, for the elucidation of the adduction site of guanine by the carcinogen retrorsine. The detection and quantification of the important biomarkers DNA adducts obtained during the reaction between carcinogen and DNA (adduction reaction) reveals the identity of the carcinogenic substances.^[1] Thus, DNA adducts analysis is complicated by their low abundance in DNA sample. This was found to range from 10 to 100 picomoles of adduct per milligram of DNA in animals, after treatment with tumorigenic doses of carcinogens.^[2] It was found that mass spectrometry is a reliable technique for the determination, characterization and sequencing of intact DNA oligomer adducts, and thus can be applied in routine analyses.^[3]

The second category of molecules that we analyzed pertained to synthetic bacterial vaccines consisting of carbohydrate haptens conjugated to the BSA, which were prepared by using the squaric acid chemistry. The development of carbohydrate glycoconjugate vaccines against bacteria is considered to be a major advance in modern medicine, since they are able to induce the eradication of the target disease.^[4]

The main scientific "raison d'être" and purpose for the analytical determination of the glycation sites of the hapten-BSA glycoconjugate vaccines are for analytical quality purposes, for checking the purity and to help to reveal the effect of the localisation of the glycation sites on the induced immune response and the production of specific well targeted antibodies. In addition, novel conjugation methods have been developed for site-specific protein modification which can be applied for the production of vaccines.^[5] The MS method that we used can thus be applied for the verification of the conjugation at specific sites of the protein in these new vaccines.

The three models were studied: the *Vibrio cholerae* O1 antigenic specific hapten conjugated to the BSA, the tetrasaccharide side chain of the *Bacillus anthracis* exosporium conjugated to the BSA, and finally the model β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside (β -D-lactoside) conjugated to the BSA.

The MALDI-TOF/TOF-MS analysis of the different synthetic carbohydrate-protein neoglycoconjugates allowed us to reveal the average carbohydrate hapten:BSA ratios.

The studies that we carried out for the glycation sites determination in hapten-protein glycoconjugate vaccines can be considered as innovative in the field of carbohydrate-protein glycoconjugate vaccines characterization. The determination of the

glycation sites of the neoglycoconjugates was attempted using different strategies. We have preferred the LC-ESI-QqTOF-MS/MS analysis to MALDI-TOF/TOF-MS/MS analysis the of the trypsin and GluC V8 digests for the sequencing of the glycated peptides. The sequencing of the glycated peptides was performed on the selected precursor ions that did not match the peptides formed by the serum albumin protein in the Mascot library.

In this thesis, it was always observed that the synthetic glycoconjugates were composed of glycoforms. This was due to the reality that the number of glycation sites was always higher than the average carbohydrate hapten:BSA ratios. For example, the glycoconjugate consisting in the conjugation of the tetrasaccharide side chain of the *Bacillus anthracis* exosporium to the BSA with a hapten:BSA ratio of 5.4 revealed 30 glycation sites. The mass spectrometry analysis of the neoglycoconjugate formed by conjugation of the antigenic monosaccharide hapten of *Vibrio cholerae* O1 serotype Ogawa to the BSA possessing a hapten:BSA ratio of 13.2:1 revealed 33 glycation sites, which was the highest number of glycation sites that we have observed amongst all the different neoglycoconjugates. Finally, a total of 30 glycation sites were also observed for the β -D-lactoside-BSA glycoconjugate vaccine model with a hapten:BSA ratio of 19.0:1. We can thus conclude that an average of 30 glycation sites appears to be the maximum number that we can determine in the glycoconjugate vaccines prepared with the squaric acid chemistry, regardless the hapten:BSA ratio.

In addition, we have observed in the GluC V8 digestion of the β -D-lactoside-BSA and the *Vibrio cholerae* O1 serotype Ogawa glycoconjugates, the formation of glycopeptides having the same peptidic sequences, which can carry either one and/or two

carbohydrate-haptens. This observation confirms the presence of glycoforms in the synthetic neoglycoconjugates. It is noteworthy to mention that these di-glycated peptides were formed only when the hapten:BSA ratio was the highest during the synthesis of each of these glycoconjugates. Therefore, we propose that this extra glycosylation of the peptide occurs despite of the possible steric hindrance imposed by the presence on the original neighboring glycosylated residue on the peptide.

It was also observed that the majority of the glycation sites in the three studied vaccines were located in the outer surface of the BSA. The importance of this discovery relies on the fact that the carbohydrate antigens are thus exposed and accessible to interact with specific receptors located at the surface of B cells.

The carrier protein BSA was successfully identified after the submission of the MS data to the Mascot library. The total sequence coverage for the hapten:BSA glycoconjugates, which is an important parameter to consider in the proteomics field, was found to be in the range of 90 % and sometimes higher. We thus concluded that our method was reliable for the mass spectrometry analysis of carbohydrate-protein glycoconjugate vaccines.

The CID-MS/MS analysis of the glycated peptides of each neoglycoconjugate also exposed product ions specific to the carbohydrate-hapten portion carried by the BSA, as well as product ions attributed to the fragmentation of the spacer $[(C_{17}H_{26}N_3O_4)^+]$ at m/z 336.19 and $(C_{11}H_{16}N_3O_2)^+$ at m/z 222.11]. It is interesting to note that the latter were observed in the three different glycoconjugates and can constitute the fingerprint of neoglycoconjugates prepared using the squaric acid chemistry.

We propose in future work to use different activation methods to initiate the fragmentation of the digested neoglycoconjugates, such as Electron Capture Dissociation (ECD with a FTICR-MS) and Electron Transfer Dissociation (ETD with an Orbitrap MS) which are known to fragment the glycopeptides only at their peptide backbone and leave the carbohydrate portion intact.

References

- [1] J. A. Miller. Recent studies on the metabolic-activation of chemical carcinogens. *Cancer Res.* **1994**, *54*, S1879-S1881.
- [2] P. S. Branco, M. P. Chiarelli, J. O. Lay Jr., F. A. Beland. Low energy tandem mass spectrometry of deoxynucleoside adducts of polycyclic aromatic hydrocarbon dihydrodiol-epoxides. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 248-256.
- [3] J. H. Banoub, S. Combden, J. Miller-Banoub, G. Sheppard, H. Hodder. Structural characterization of intact covalently linked DNA adducts by electrospray mass spectrometry. *Nucleosides & Nucleotides.* **1999**, *18*, 2751-2768.
- [4] R. A. Pon. Exploiting the bacterial surface: the successful application of glycoconjugate vaccines. In *Bacterial Glycomics: Current Research, Technology and Applications*. C. W. Reid, S. M. Twine, A. N. Reid (Eds.). *Caister Academic Press* (**2012**), pp. 257.
- [5] W. Ou, T. Uno, H. P. Chiu, J. Grünewald, S. E. Cellitti, T. Crossgrove, X. Hao, Q. Fan, L. L. Quinn, P. Patterson, L. Okach, D. H. Jones, S. A. Lesley, A. Brock, B. H. Geierstanger. Site-specific protein modifications through pyrroline-carboxy-lysine residues. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10437-10442.

APPENDIX A

Additional Figures and tables for

CHAPTER 2: Differentiation and sequencing of three constitutional isobaric 18-mer
DNA oligomers using low-energy collision tandem mass spectrometry.

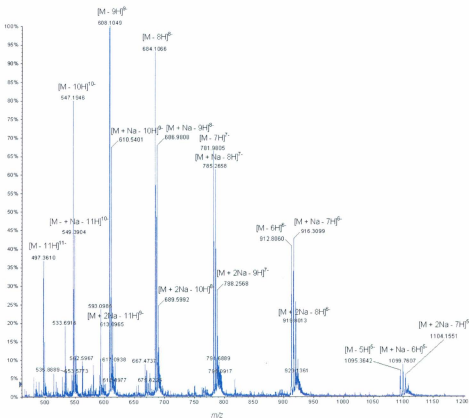


Figure A.1. ESI-QqTOF MS (negative mode) of the oligomer 1.

Table A.1. Different specific product ions of the oligomer **1** obtained during the MS/MS analysis of the $[M - 8H]^{8-}$ anion at m/z 683.9881 and the $[M - 9H]^{9-}$ anion at m/z 607.8775.

Ion	Calculated m/z	Deviation for $[M - 8H]^{8-}$ (ppm)	Deviation for $[M - 9H]^{9-}$ (ppm)
$[d_1 - H_2O]^-$	328.0016	50	9
$[G_9;C_{10}]^{2-}$	397.0420	-6	-4
w_4^{3-}	411.3970	-	-7
w_6^{4-}	462.8200	-	-4
w_5^{3-}	521.0810	-	-8
$[T_3;C_5]^{2-}$	536.5610	-	-16
$[a_6 - B]^{3-}$	544.7530	4	-23
$[G_9;T_{11}]^{2-}$	549.0650	-	-6
$[G_9;C_{13}]^{3-}$	566.4080	-	1
$[T_7;T_{11}]^{3-}$	571.4080	-	-16
$[C_5;G_9]^{3-}$ $[A_6;C_{10}]^{3-}$ $[T_7;T_{11}]^{3-}$ $[A_8;A_{12}]^{3-}$ $[T_{11};A_{15}]^{3-}$ $[C_{13};T_{17}]^{3-}$	574.4120	-2	-29
y_4^{2-}	577.6170	-	23
w_8^{4-}	617.0960	-7	-
w_6^{3-}	617.4300	-	19
$[a_7 - B]^{3-}$	649.1050	-4	-13
$[M - A]^{8-}$	667.2300	-21	-22
$[M - CH]^{8-}$	670.1060	-	-23
$[a_5 - B]^{2-}$	673.1100	-	-13
$[G_9;G_{14}]^{3-}$	676.0920	-	-9
$[a_{12} - B]^{3-}$	696.9090	52	-15
y_{10}^{4-}	751.6290	-	-16
$[d_5 - H_2O]^{2-}$	768.6150	-3	-12
$[a_{16} - B]^{6-}$	787.9580	-1	-1
$[A_{15};A_{16}]^-$	803.1080	-11	-
$[a_6 - B]^{2-}$	817.6330	-	-16
$[a_{12} - B]^{4-}$	871.3890	-	-16
$[d_6 - H_2O]^{2-}$	925.1440	-	-15
w_6^{2-}	926.6490	-	-14

$[d_3 - H_2O]^-$	945.1470	-3	-9
$[a_7 - B]^{2-}$	974.1620	-	-25
$[T_4; G_9]^{2-}$	1014.1450	-5	-
$[A_6; T_{11}]^{2-}$			
$[T_7; A_{12}]^{2-}$			
$[A_6; A_8]^-$	1107.1540	10	-9
$[A_{15}; T_{17}]^-$			
$[a_8 - B]^{2-}$	1126.1850	-	2

Table A.2. Different specific product ions of the oligomer **2** obtained during the MS/MS analysis of the $[M - 8H]^{8-}$ anion at m/z 683.9881 and the $[M - 9H]^{9-}$ anion at m/z 607.8775.

Ion	Calculated m/z	Deviation for $[M - 8H]^{8-}$ (ppm)	Deviation for $[M - 9H]^{9-}$ (ppm)
$[T_4:C_5]^{2-}$ $[C_{10}:T_{11}]^{2-}$	384.539	0	-
$[T_3:T_4]^{2-}$	392.038	-21	-
$[A_2:T_3]^{2-}$ $[T_7:A_8]^{2-}$ $[T_{11}:A_{12}]^{2-}$ $[A_{16}:T_{17}]^{2-}$	396.544	12	-
$[C_5:G_6]^{2-}$ $[G_9:C_{10}]^{2-}$ $[C_{13}:G_{14}]^{2-}$	397.042	4	-
$[G_6:T_7]^{2-}$	404.542	2	-
y_2^+	530.127	-15	1
$[A_2:T_4]^{2-}$	548.567	-	-1
$[G_6:A_8]^{2-}$ $[T_7:G_9]^{2-}$	561.07	-	-49
$[G_{14}:A_{16}]^{2-}$	565.576	4	-
$[G_9:C_{13}]^{3-}$	566.408	-3	-
$[A_8:A_{12}]^{3-}$ $[T_{11}:A_{15}]^{3-}$ $[C_{13}:T_{17}]^{3-}$	574.412	-1	-
$[d_4 - H_2O]^{2-}$	616.095	2	-
$[M - 8H]^{8-}$	665.105	-	-9
$[A_2:T_7]^{3-}$ $[T_3:A_8]^{3-}$	672.757	-6	-
$[C_5:C_{10}]^{3-}$ $[G_9:G_{14}]^{3-}$	676.092	1	-
$[C_{10}:C_{13}]^{2-}$	685.59	8	-
y_7^{3-}	695.127	-28	-

$[T_3; G_6]^{2-}$ $[T_4; T_7]^{2-}$	701.088	9	-
$[a_3 - B]^-$	723.142	-9	-
$[G_6; G_9]^{2-}$	725.596	-	-3
y_7^{2-}	1043.195	-14	-
$[G_{14}; A_{16}]^-$	1132.161	13	17

Table A.3. Different specific product ions of the oligomer **3** obtained during the MS/MS analysis of the $[M - 8H]^{8-}$ anion at m/z 683.9881 and the $[M - 9H]^{9-}$ anion at m/z 607.8775.

Ion	Calculated m/z	Deviation for $[M - 8H]^{8-}$ (ppm)	Deviation for $[M - 9H]^{9-}$ (ppm)
$[T_3;T_4]^{2-}$	392.0380	-	11
$[T_3;T_7]^{+}$	426.0510	-	-31
w_6^{4-}	472.8220	33	-
w_7^{4-}	551.0860	5	-
$[G_6;A_8]^{2-}$	561.0700	2	-
$[a_8 - B]^{+}$	562.5880	13	-
w_{12}^{6-}	616.9280	73	-
w_{10}^{5-}	617.0950	-	-26
$[G_6;C_{13}]^{+}$ $[T_7;G_{14}]^{+}$	651.0910	17	17
w_{15}^{7-}	660.3860	23	-
$[M - AH]^{8-}$	667.1040	3	-
$[a_{15} - B]^{6-}$	729.1140	-30	-
$[M - A]^{7-}$	762.6930	-15	-
$[T_4;C_{10}]^{3-}$ $[C_5;T_{11}]^{3-}$	764.1050	-23	-
w_{10}^{4-}	771.6210	12	-
$[A_5;A_8]^{1-}$	777.1090	14	-
$[A_{12};C_{13}]^{-}$	779.0970	-	-36
$[A_2;T_{11}]^{+}$ $[T_3;A_{12}]^{+}$	803.1140	-	-1
y_9^{3-}	906.1600	-	-7
$[G_6;T_{11}]^{2-}$ $[T_4;C_9]^{2-}$	1002.1390	-28	-
$[C_9;T_{11}]^{-}$	1059.1320	-14	-
$[A_{12};G_{14}]^{-}$ $[C_{13};A_{15}]^{-}$	1108.1490	12	-

Table A.4. Product ions common (isobaries and non isobaries) to the three oligomers obtained during the MS/MS analysis of the $[M - 8H]^{8-}$ anion at m/z 683.9881.

Ion	Oligomer 1		Oligomer 2		Oligomer 3	
	Experimental mass (Daltons)	Deviation (ppm)	Experimental Mass (Daltons)	Deviation (ppm)	Experimental mass (Daltons)	Deviation (ppm)
$[pd(T)]^-$	303.0392	-1	303.0396	-2	303.0470	-27
w_2^{2-}	304.5429	0	304.5430	0	324.5523	19
w_1^-	306.0488	3	306.0479	0	346.0610	20
$[pd(A)]^-$	312.0491	4	312.0490	4	312.0586	-27
$[a_2 - B]^-$	426.0820	5	410.0818	-8	410.0853	1
w_3^{2-}	461.0687	-5	461.0684	-6	481.0689	-11
$[pd(C)pf]^-$	466.0367	-7	466.0402	0	466.0437	8
$[pd(T)pf]^-$	481.0406	1	481.0410	2	-	-
$[pd(A)pf]^-$	490.0512	0	490.0507	-1	490.0594	17
$[pd(G)pf]^-$	506.0459	0	506.0474	3	506.0498	8
w_5^{3-}	521.0813	1	521.0792	-3	534.4272	19
$[T_3:C_5]^{2-}$	536.5706	18	536.5631	4	-	-
$[C_5:T_7]^{2-}$	541.0725	10	549.0629	-4	549.0675	5
$[C_{13}:A_{15}]^{2-}$ $[A_{12}:G_{14}]^{2-}$	553.5673	-7	553.5689	-4	553.5813	19
$[T_7:G_9]^{2-}$	561.0526	-31	561.0698	0	-	-
$[C_{10}:G_{14}]^{3-}$	-	-	568.4007	-13	566.4138	10
$[T_7:T_{11}]^{3-}$	571.4004	-13	571.4099	3	-	-
y_4^{2-}	577.6099	-12	577.6193	4	-	-
w_2^-	610.0910	-3	610.0934	1	650.0968	-5
w_6^{3-}	617.4181	-19	617.4261	-6	630.7709	9
$[d_2 - H_2O]^-$	641.0936	-12	625.0860	-32	-	-
$[M - CH]^{8-}$	670.0987	-11	670.0938	-18	-	-
$[a_5 - B]^{2-}$	673.1054	-7	665.1079	-8	665.1061	-10
$[T_4:G_9]^{1-}$	675.7594	-1	681.0911	-1	-	-
$[T_{11}:G_{14}]^{2-}$	705.5935	1	705.5939	1	705.5994	9

$[C_{13};A_{16}]^{2-}$ $[A_{12};A_{15}]^{2-}$	710.0807	-26	710.0962	-4	-	-
$[a_3 - B]^-$	739.1342	-4	723.1356	-9	-	-
y_5^{2-}	742.1333	-13	742.1423	-1	-	-
w_{17}^{7-}	-	-	748.5498	8	748.5771	44
$[a_8 - B]^{3+}$	750.4478	-8	750.4530	-1	750.4395	-19
$[C_{10};T_{11}]^-$ $[T_4;C_5]^-$	770.0776	-11	770.0856	-1	770.0639	-29
w_5^{2-}	782.1231	-4	782.1239	-3	802.1492	25
$[T_3;T_4]^-$	785.0812	-5	785.0850	0	785.1037	24
$[a_{16} - B]^{6+}$	787.9272	-39	-	-	781.2867	-6
$[A_2;T_3]^-$ $[T_7;A_8]^-$ $[T_{11};A_{12}]^-$ $[A_{16};T_{17}]^-$	794.0972	0	794.1016	6	794.1041	9
$[C_{13};G_{14}]^-$	795.0934	2	795.0970	6	795.0943	3
$[a_6 - B]^{2+}$	817.6290	-5	809.6368	1	-	-
$[G_{14};A_{15}]^-$	819.1000	-4	819.1023	-1	-	-
y_3^-	843.1699	-18	843.1886	4	883.1763	-17
$[T_3;T_7]^{2+}$	845.1041	-11	853.1149	6	-	-
$[C_{10};G_{14}]^{2+}$	850.1224	8	850.1313	18	-	-
$[T_7;T_{11}]^{2+}$	857.6094	-8	857.6186	3	837.6124	-1
$[A_8;A_{12}]^{2-}$ $[T_{11};A_{15}]^{2+}$	862.1096	-27	862.1236	2	-	-
$[a_{12} - B]^{4+}$	871.3856	-4	-	-	861.3868	0
y_6^{2-}	886.6554	-12	886.6712	6	-	-
w_3^-	923.1469	-4	923.1490	-2	963.1609	4
w_6^{2-}	926.6393	-10	926.6501	1	-	-
$[a_7 - B]^{2+}$	974.1417	-21	974.1552	-7	-	-
$[A_2;T_7]^{2+}$ $[T_3;A_8]^{2+}$	1001.6314	-11	1009.6401	4	-	-
$[a_4 - B]^-$	1043.1860	3	1027.1830	-4	-	-
$[T_3;C_5]^-$	1074.1331	2	1074.1372	6	1074.1335	2

$[C_{10}:A_{12}]^-$ $[T_{11}:C_{13}]^-$	1083.1411	-2	1083.1433	0	1083.1198	-21
$[A_2:T_4]^-$	1098.1627	18	1098.1458	3	1098.1663	21
$[C_{13}:A_{15}]^-$ $[A_{12}:G_{14}]^-$	-	-	1108.1528	3	1108.1628	12
$[T_7:G_9]^-$	1123.1372	-11	1123.1597	10	-	-
$[a_8 - B]^{2-}$	1126.1601	-22	1126.1752	-9	-	-
y_4^-	1156.2282	-12	1156.2388	-3	-	-

Table A.5. Product ions common (isobaries and non isobaries) to the three oligomers obtained during the MS/MS analysis of the $[M - 9H]^{9-}$ anion at m/z 607.8775.

Ion	Oligomer 1		Oligomer 2		Oligomer 3	
	Experimental mass (Daltons)	Deviation (ppm)	Experimental Mass (Daltons)	Deviation (ppm)	Experimental mass (Daltons)	Deviation (ppm)
$[pd(T)]^-$	303.0392	-1	303.0389	0	303.0364	8
w_2^{2-}	304.5428	-1	304.5426	-1	324.5525	20
w_1^-	306.0502	7	306.0476	-1	346.0505	-10
$[pd(A)]^-$	312.0483	6	312.0498	2	312.0489	4
$[a_2 - B]^-$	426.0780	-5	410.0795	-13	-	-
w_3^{2-}	461.0683	-6	461.0740	7	-	-
$[pd(C)pf]^-$	466.0372	-6	466.0410	2	466.0392	-2
$[pd(T)pf]^-$	481.0394	-1	481.0434	7	481.0419	4
$[pd(A)pf]^-$	490.0458	-11	490.0536	5	490.0536	5
$[pd(G)pf]^-$	506.0415	-9	506.0486	5	506.0497	7
$[C_5:T_7]^{2-}$	541.0697	5	549.0676	5	-	-
w_2^-	610.0859	-12	610.0950	3	650.0981	-3
w_8^{4-}	617.0860	-16	617.0913	-8	-	-
$[d_2 + H_2O]^-$	641.0889	-19	625.1073	2	625.0925	-22
$[T_4:T_7]^{2-}$	693.0824	-11	701.0920	6	-	-
$[T_{11}:G_{14}]^{2-}$	705.5829	-14	705.5988	8	705.5646	-40
$[A_{12}:A_{13}]^{2-}$ $[C_{13}:A_{16}]^{2-}$	710.0799	-27	710.0987	0	-	-
$[a_3 - B]^-$	739.1292	-11	723.1482	9	723.1172	-34
y_5^{2-}	742.1295	-18	742.1401	-4	-	-
$[T_4:C_5]^-$ $[C_{10}:T_{11}]^-$	770.0816	-6	770.0897	5	770.0782	-10
w_5^{2-}	782.1193	-9	782.1170	-12	-	-
$[T_3:T_4]^-$	785.0712	-18	785.0929	10	785.0917	9
$[A_5:T_3]^-$	794.0880	-11	794.0977	1	794.1093	15
$[T_7:A_8]^-$						
$[T_{11}:A_{12}]^-$						

[A ₁₆ :T ₁₇]						
[C ₁₃ :G ₁₄]	795.0903	-2	795.0998	10	795.1089	21
[A ₁₅ :A ₁₆]	803.1005	-9	803.1113	4	-	-
[G ₆ :T ₇]	-	-	810.1019	12	810.1115	24
[G ₁₄ :A ₁₅]	819.1042	1	819.1064	4	819.0710	-39
y ₃ ⁻	843.1642	-25	843.1850	0	883.1971	7
[T ₃ :T ₇] ²⁻	845.1009	-14	853.1120	2	-	-
[C ₁₀ :G ₁₄] ²⁻	850.1005	-18	850.0952	-24	-	-
[T ₇ :T ₁₁] ²⁻	857.6033	-15	857.6394	27	-	-
[A ₈ :A ₁₂] ²⁻	862.1038	-21	862.1391	20	-	-
[T ₁₁ :A ₁₅] ²⁻						
y ₆ ²⁻	886.6485	-20	886.6599	-7	-	-
w ₃ ⁻	923.1414	-10	923.1549	4	963.1358	-22
[a ₄ - B]	1043.1532	-29	1027.1883	0	-	-
[T ₃ :C ₅]	1074.1156	-14	1074.1291	-2	1074.1153	-15
[C ₁₀ :A ₁₂]	1083.1379	-5	1083.1436	1	1083.1611	17
[A ₂ :T ₄]	1098.1356	-7	1098.1530	9	-	-
[G ₉ :T ₁₁]	1099.1149	-18	1099.1452	7	-	-
[A ₈ :C ₁₀]	1108.1247	-22	1108.1650	14	-	-
[A ₁₂ :G ₁₄]						
[C ₁₃ :A ₁₅]						
[T ₇ :G ₉]	1123.1278	-19	1123.1557	6	-	-
y ₄ ⁻	1156.2161	-22	1156.2549	11	-	-

APPENDIX B

Additional Figures and Tables for

**CHAPTER 3: Glycation sites in neoglycoconjugates from the terminal
monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and
BSA revealed by matrix-assisted laser desorption-ionization tandem mass
spectrometry**

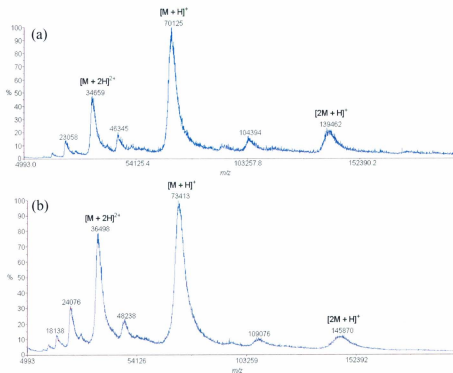


Figure B.1. MALDI-TOF/TOF-MS analysis of the hapten-BSA neoglycoconjugates with a hapten:protein ratio of 6.6 (a) and 13.2 (b).

Table B.1. MALDI-TOF/TOF-MS/MS of the glycanated peptide SLGK*VGTR at m/z 1330.72.

Ion	Observed m/z	Calculated m/z	Deviation (Da)
$[M + H]^+$	1332.4247	1330.7213	1.7034
$[M + H - CH_4]^+$	1314.6834	1314.69	-0.0066
$[M + H - C_{11}H_{19}NO_6]^+$	1069.5933	1069.6	-0.0067
$[M + H - C_{11}H_{19}NO_6 - NH_3]^+$	1052.5676	1052.5623	0.0053
$[y_5 - CH_2O_2]^+$	1027.5660	1027.5782	-0.0122
$[z_6 - C_{11}H_{19}NO_6]^+$	852.4828	852.4579	0.0249
$[b_5 - C_{11}H_{19}NO_6]^+$	737.4564	737.4198	0.0366
$[b_4 - C_{11}H_{19}NO_6]^+$	638.3859	638.3513	0.0346
$[a_4 - C_{15}H_{28}NO_7]^+$	537.3314	537.2905	0.0409
$[K*VGTR - C_{10}H_{14}N_2O_8]^+$	438.2656	438.2347	0.0309
y_4^+	432.2989	432.2565	0.0424
$C_{13}H_{30}NO_7^+$	336.2110	336.2017	0.0093
y_3^+	333.2013	333.1881	0.0132
y_2^+	276.1725	276.1666	0.0059
$C_{11}H_{20}NO_6^+$	262.1366	262.1290	0.0076
$C_{11}H_{18}NO_5^+$	244.1309	244.1179	0.0130
y_1^+	175.1199	175.1190	0.0009
$C_6H_8NO_3^+$	142.0679	142.0499	0.0180
$[y_1 - CH_2O_2]^+$	129.1205	129.1135	0.0070
$[z_1 - CH_2O_2]^+$	112.0758	112.0869	-0.0111

Table B.2. MALDI-TOF/TOF-MS/MS of the glycosylated peptide ALK*AWSVAR at m/z 1514.82.

Ion	Observed m/z	Calculated m/z	Deviation (Da)
$[M + H]^+$	1517.8384	1514.8213	3.0171
$[M + H - C_3H_6O_2]^+$	1440.0994	1440.7845	-0.6851
$[M + H - C_{11}H_{19}NO_6]^+$	1253.6884	1253.7001	-0.0117
$[M + H - C_{11}H_{20}NO_7]^-$	1236.6424	1236.6973	-0.0549
$[M + H - C_{19}H_{36}N_2O_8]^+$	1092.5640	1092.5585	0.0055
$[c_7 - C_{17}H_{31}NO_8]^+$	909.5271	909.4941	0.0330
$[b_5 - C_{11}H_{19}NO_6]^+$	822.5058	822.4514	0.0544
$[a_5 - C_{11}H_{19}NO_6]^+$	794.4930	794.4565	0.0365
$[a_6 - C_{19}H_{35}N_2O_8]^-$	723.4375	723.3704	0.0671
y_6^+	689.3879	689.3729	0.0150
$[a_5 - C_{19}H_{35}N_2O_8]^+$	636.4226	636.3384	0.0842
y_5^+	618.3445	618.3358	0.0087
$[b_3 - C_{11}H_{19}NO_6]^+$	565.3745	565.3350	0.0395
y_4^+	432.2845	432.2565	0.0280
y_3^+	345.2283	345.2245	0.0038
$C_{15}H_{30}NO_7^+$	336.2042	336.2017	0.0025
$C_{11}H_{20}NO_6^+$	262.1406	262.1290	0.0116
y_2^+	246.1630	246.1561	0.0069
$C_{11}H_{18}NO_5^+$	244.1282	244.1179	0.0103
y_1^+	175.1198	175.1190	0.0008
$[y_1 - CO_2]^+$	129.1129	129.1135	-0.0006
$[z_1 - CH_2O_2]^+$	112.0862	112.0869	-0.0007

Table B.3. MALDI-TOF/TOF-MS/MS of the glycosylated peptide K*VPQVSTPTLVEVSR at m/z 2153.16.

Ion	Observed m/z	Calculated m/z	Deviation (Da)
$[M + H]^+$	2153.6801	2153.1700	0.5101
$[M + H - C_2H_5O]^+$	2108.1530	2107.1281	1.0249
$[M + H - C_{11}H_{19}NO_6]^+$	1892.0073	1892.0487	-0.0414
y_{13}^+	1412.7814	1412.7744	0.0070
z_{13}^+	1396.8246	1396.7562	0.0684
y_{11}^+	1187.7466	1187.6630	0.0836
y_{10}^+	1088.6348	1088.5946	0.0402
y_9^+	1001.6400	1001.5626	0.0774
y_8^+	900.5673	900.5149	0.0524
z_8^+	884.5682	884.4967	0.0715
$[b_5 - C_7H_{13}NO_3]^+$	804.5096	804.4620	0.0476
y_7^+	803.5045	803.4621	0.0424
$[b_4 - C_{11}H_{19}NO_6]^+$	705.4496	705.3935	0.0561
y_6^+	702.4599	702.4145	0.0454
y_5^+	589.3736	589.3304	0.0432
$C_{23}H_{35}N_3O_{10}^{+}$	513.3080	513.2317	0.0763
$C_{23}H_{33}N_3O_9^{+}$	495.2560	495.2211	0.0349
y_4^+	490.2816	490.2620	0.0196
$[b_2 - C_{11}H_{19}NO_6]^+$	480.3265	480.2817	0.0448
$[b_4 - C_{23}H_{35}N_3O_{10}]^{+}$	452.3134	452.2742	0.0392
$C_{19}H_{30}N_3O_7^{+}$	412.2450	412.2078	0.0372
y_3^+	361.2359	361.2194	0.0165
$C_{15}H_{30}NO_7^{+}$	336.2131	336.2017	0.0114
$[b_3 - C_{23}H_{35}N_3O_{10}]^{+}$	325.2255	325.2234	0.0021
$C_{15}H_{26}N_3O_4^{+}$	312.2119	312.1918	0.0201
$[a_3 - C_{23}H_{35}N_3O_{10}]^{+}$	297.2130	297.2285	-0.0155
y_2^+	262.1570	262.1510	0.0060
$C_{11}H_{18}NO_5^{+}$	244.1331	244.1179	0.0152
$[b_2 - C_{23}H_{37}N_3O_{10}]^{+}$	226.1504	226.1556	-0.0052
y_1^+	175.1199	175.1190	0.0009
$[y_1 - CH_2O_2]^+$	129.1108	129.1135	-0.0027
$[z_1 - CH_2O_2]^+$	112.0833	112.0869	-0.0036

APPENDIX C

Additional Figures and Tables for

**CHAPTER 4: Determination of the glycation sites of *Bacillus anthracis*
neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-
QqTOF-tandem mass spectrometry**

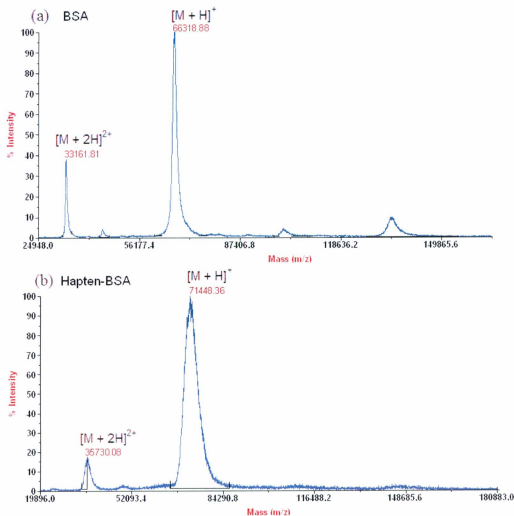


Figure C.1. MALDI-TOF/TOF-MS analysis of (a) the BSA and (b) the hapten-BSA glycoconjugate.

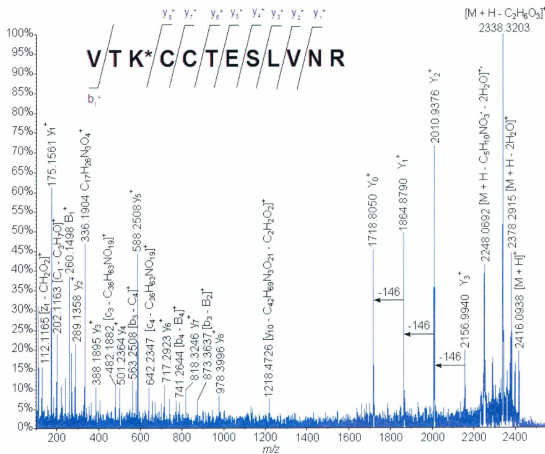


Figure C.2. MALDI-TOF/TOF-MS/MS spectra of the glycosylated peptide VTK*CCTESLVNR (Lys 498) at m/z 2416.1372.

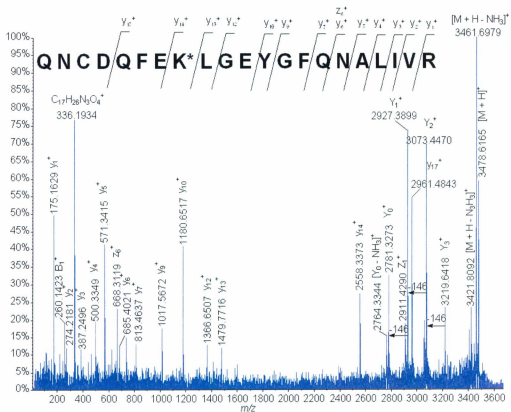


Figure C.3. MALDI-TOF/TOF-MS/MS spectra of the glycated peptide QNCDQFEK*LGEYGFQNALIVR (Lys 420) at m/z 3478.6429.

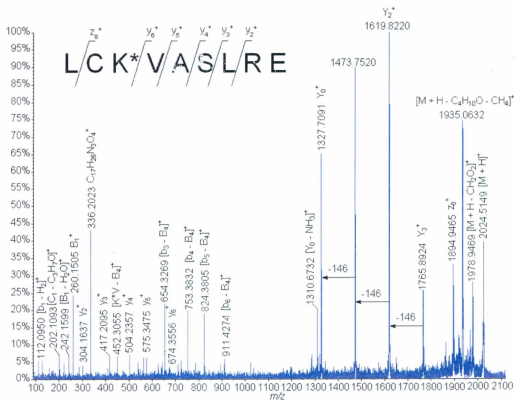


Figure C.4. MALDI-TOF/TOF-MS/MS spectra of the glycosylated peptide LCK*VASLRE (Lys 100) at m/z 2025.0186.

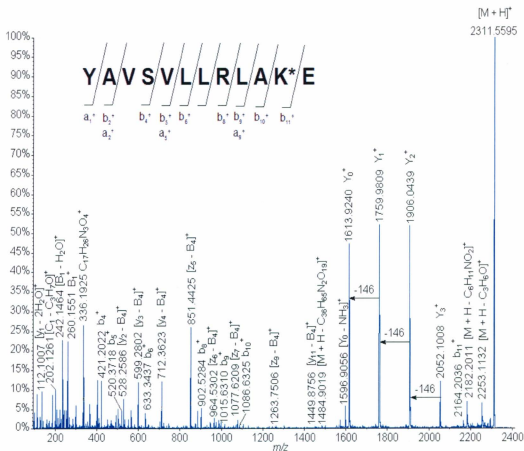


Figure C.5. MALDI-TOF/TOF-MS/MS spectra of the glycosylated peptide YAVSVLLRLAK*E (Lys 374) at m/z 2311.2476.

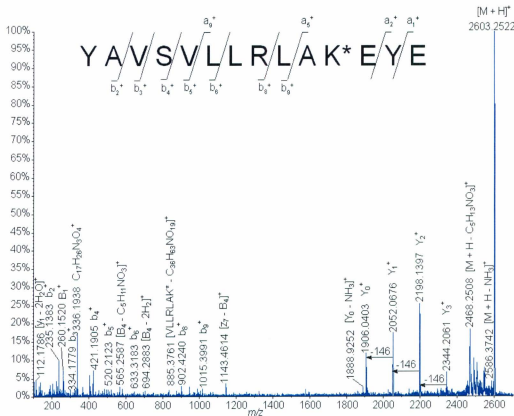


Figure C.6. MALDI-TOF/TOF-MS/MS spectra of the glycosylated peptide YAVSVLLRLAK*EYE (Lys 374) at m/z 2603.3484.

Table C.1. Tryptic peptides identified in the two serum albumin protein isoforms by MALDI-TOF/TOF-MS analysis of the hapten-BSA glycoconjugate. The experimental values were obtained after recalibration of the spectra.

Observed <i>m/z</i>	Calculated <i>m/z</i>	Accuracy ppm	Missed cleavage	Serum albumin <i>Bos taurus</i> (gi 74267962)	Serum albumin precursor <i>Bos taurus</i> (gi 1351907)	Unique	Peptide
847.5055	847.5041	-2	1	✓	✓		LSQKFPK
898.4808	898.4820	1	0	✓	✓		LCVLHEK
922.4853	922.4885	4	0	✓	✓		AEFVEVTK
927.4940	927.4940	0	0	✓	✓		YLYEiar
1014.6208	1014.6199	-1	0	✓	✓		QTALVELLK
1138.5020	1138.4985	-3	0	✓	✓		CCTESLVNR
1142.7170	1142.7149	-2	1	✓	✓		KQTALVELLK
1163.6336	1163.6312	-2	0	✓	✓		LVNELTEFAK
1193.6064	1193.6027	-3	1	✓	✓		DTHKSEIAHR
1249.6228	1249.6217	-1	1	✓	✓		FKDLGEEHFK
1283.7148	1283.7112	-3	0	✓	✓		HPEYAVSVLLR
1305.7177	1305.7167	-1	0	✓	✓		HLVDEPQNLIK
1419.6945	1419.6942	0	0	✓	✓		SLHTLFGDELCK
1439.8139	1439.8123	-1	1	✓	✓		RHPEYAVSVLLR
1479.7977	1479.7960	-1	0	-	✓		LGEYGFQNALIVR
1532.7783	1532.7816	2	1	✓	✓		LKECCDKPLLEK
1537.8041	1537.8014	-2	0	✓	-	U	LGEYGFQNELIVR
1554.6639	1554.6535	-7	0	✓	✓		DDPHACYSTVFDDK
1567.7471	1567.7433	-2	0	✓	✓		DAFLGSFLYEYSR

1576.7635	1576.7681	3	0	√	√		LKPDPNTLCDEFK
1639.9417	1639.9383	-2	1	√	√		KVPQVSTPTLVEVSR
1661.9224	1661.8718	-30	2	√	√		GACLLPKIETMREK
1694.8245	1694.8246	0	0	√	-	U	MPCAEDYLSLILNR
1724.8353	1724.8351	0	0	-	√		MPCTEDYLSLILNR
1880.9275	1880.9216	-3	0	√	√		RPCFSALTPDETYVPK
1901.8758	1901.8703	-3	1	√	√		NECFLSHKDDSPDLPK
1907.9231	1907.9213	-1	0	√	√		LFTFHADICTLPDTEK
1927.8001	1927.7988	-1	1	√	√		CCAADDKEACFAVEGPK
1951.0130	1951.0158	1	1	√	√		ALK*AWSVAR (Lys 235)
2019.9644	2019.9697	3	1	√	√		LKPDPNTLCDEFKADEK
2045.0307	2045.0285	-1	1	√	√		RHPYFYAPPELLYYANK
2113.8960	2113.8853	-5	1	√	√		VHKECCHGDLLECADDR
2247.9470	2247.9433	-2	1	√	√		ECCHGDLLECADDRADLAK
2416.1372	2416.1357	-1	1	√	√		VTK*CTESLVNR (Lys 498)
2458.1766	2458.1812	2	1	√	√		DAIPENLPPLTADFAEDKDVCK
2492.2653	2492.2648	0	0	√	√		GLVLIAFSQYLQCPFDEHVK
2541.1679	2541.1680	0	2	√	√		QEPERNECFLSHKDDSPDLPK
2612.1872	2612.1655	-8	2	√	√		VHKECCHGDLLECADDRADLAK
3250.5163	3250.4681	-15	2	√	-		TVMENFVAFVGKCCAADDKEACFAVEGPK
3478.6429	3478.6458	1	1	-	√		QNCDQFEK*LGEYGFQNALIVR (Lys 420)
3511.6771	3511.6725	-1	2	√	√		SHCIAEVEKDAIPENLPPLTADFAEDKDVCK

Table C.2. Peptides identified in the two serum albumin protein isoforms by MALDI-TOF/TOF-MS analysis of the digested hapten-BSA glycoconjugate with the endoproteinase GluC V8. The experimental values were obtained after recalibration of the spectra.

Observed <i>m/z</i>	Calculated <i>m/z</i>	Accuracy ppm	Missed cleavage	Serum albumin <i>Bos taurus</i> (gi 74267962)	Serum albumin precursor <i>Bos taurus</i> (gi 1351907)	Unique	Peptide
838.4787	838.4787	0	0	✓	✓		HVKLVNE
844.4019	844.3987	-4	0	✓	✓		KSHCIAE
848.4568	848.346	-131	1	✓	✓		PNTLCDE
854.5126	854.51	-3	0	✓	✓		LTKVHKE
944.4718	944.4702	-2	0	✓	✓		YSRRHPE
978.505	978.5049	0	0	✓	✓		KLFTFHAD
1003.3963	1003.3977	1	1	✓	✓		CCHGDLLE
1017.4364	1017.5369	99	0	✓	✓		KSLHTLFGD
1025.4248	1025.4032	50	0	✓	✓		KVTKCCTE
1029.4978	1029.4965	-1	2	-	✓		KQEPERNE
1034.4639	1034.4651	1	1	✓	✓		CCDKPLLE
1040.4626	1040.4722	9	1	✓	✓		FAKTCVADE
1069.5166	1069.5206	4	1	✓	✓		ETYVPKAFD
	1069.5206	4	1	✓	✓		TYVPKAFDE
1075.5958	1075.5934	-2	0	✓	✓		LCKVASLRE
1146.5792	1146.5795	0	1	✓	✓		KSLHTLFGDE
1243.6739	1243.672	-2	1	✓	✓		DKGACLLPKIE
1284.7502	1284.7527	2	0	✓	✓		GPKLVVSTQTALA
1293.7541	1293.753	-1	1	✓	✓		LLKHKPKATEE
1314.6833	1314.6806	-2	2	✓	✓		IAHRFKDLGEE
1315.7303	1315.6316	-75	1	✓	✓		KSHCIAEVEKD
1360.6739	1360.6789	4	1	✓	✓		AKDAFLGSFLYE
1383.6659	1383.6466	-14	2	✓	✓		LTEFAKTCVADE
1519.7811	1519.781	0	0	✓	✓		IARRHPYFYAPE

1653.9224	1653.9216	0	1	√	√		YAVSVLLRLAKEYE
1721.8452	1721.8539	5	0	√	√		LLYYANKYNGVFQE
1742.8396	1742.9627	71	0	√	√		YLSLILNRLCVLHE
1857.9968	1857.9897	-4	1	√	√		DYLSLILNRLCVLHE
1907.9236	1907.9213	-1	2	√	√		KLFTFHADICTLPDTE
1934.9962	1934.9904	-3	0	√	√		VSRSLGKVGTRCCTKPE
1962.0323	1962.037	2	1	√	√		ACFAVEGPKLVVSTQTALA
2025.0186	2025.0195	0	0	√	√		LCK*VASLRE (Lys 100)
2311.2476	2311.2418	-3	0	√	√		YAVSVLLRLAK*E (Lys 374)
2467.1422	2467.1451	1	4	√	√		NLPPLTADFAEDKDVCKNYQE
2603.3484	2603.3477	0	1	√	√		YAVSVLLRLAK*EYE (Lys 374)
2879.5814	2879.5783	-1	0	-	√	U	YGFQNALIVRYTRKVPQVSTPTLVE
2912.4256	2912.4252	0	3	√	√		SLVNRRPCFSALTPDETYVPKAFDE
2941.348	2941.3752	9	2	-			VSRSLGKVGTRCCTKPESERMPCTE

Table C.3. MALDI-TOF/TOF-MS/MS analysis of the glycosylated peptide ALK*AWSVAR (Lys 235) at m/z 1951.0130.

Molecular ion	Calculated m/z	Experimental m/z	Deviation (Da)
$[M + H]^+$	1951.0158	1951.0180	0.0022
$[M + H - CH_2O]^+$	1921.0052	1921.0932	0.0880
$[M + H - CH_{10}O_3]^+$	1880.9528	1881.1765	0.2237
Y_3^+	1691.8738	1691.8829	0.0091
Y_2^+	1545.8159	1545.8376	0.0217
Y_1^+	1399.758	1399.7768	0.0188
$[y_7 - B_2]^+$	1361.6947	1361.7265	0.0318
Y_0^+	1253.7001	1253.7010	0.0009
$[Y_0 - NH_3]^+$	1236.6735	1236.6813	0.0078
$[b_3 - C_3H_{11}NO_2]^+$	1145.5711	1145.5757	0.0046
$[b_4 - B_4]^+$	636.3715	636.2715	-0.1000
y_5^+	618.3558	618.2805	-0.0753
$[b_3 - B_4]^+$	565.3442	565.2661	-0.0781
y_3^+	345.2245	345.1345	-0.0900
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1242	-0.0676
B_1^+	260.1492	260.0739	-0.0753
y_2^+	246.1561	246.0775	-0.0786
$[B_1 - H_2O]^+$	242.1387	242.0709	-0.0678
$^{2,5}A_1^+$	230.1387	230.0719	-0.0668
z_2^+	229.1306	229.0569	-0.0737
$C_{11}H_{16}N_3O_2^+$	222.1237	222.0670	-0.0567
$[C_1 - C_3H_5O]^+$	202.1074	202.0350	-0.0724
y_1^+	175.119	175.0651	-0.0539
$[C_1 - C_3H_{11}NO_2]^+$	159.0652	159.0486	-0.0166
$[y_1 - CH_2O_2]^+$	129.1134	129.0679	-0.0455
$[z_1 - CH_2O_2]^+$	112.0869	112.0527	-0.0342
$[z_1 - CH_4O_2]^+$	110.0713	110.0414	-0.0299

Table C.4. MALDI-TOF/TOF-MS/MS analysis of the glycated peptide LCK*VASLRE (Lys 100) at m/z 2025.0186.

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
$[M + H]^+$	2025.0195	2024.5149	-0.5046
$[M + H - CH_2O_2]^+$	1979.0141	1978.9469	-0.0672
$[M + H - C_4H_{10}O - CH_4]^+$	1934.9151	1935.0632	0.1481
z_8^+	1894.9089	1894.9465	0.0376
Y_3^+	1765.8776	1765.8924	0.0148
Y_2^+	1619.8197	1619.8220	0.0023
Y_1^+	1473.7618	1473.7520	-0.0098
Y_0^+	1327.7038	1327.7091	0.0053
$[Y_0 - NH_3]^+$	1310.6773	1310.6732	-0.0041
$[b_6 - B_4]^+$	911.4661	911.4274	-0.0387
$[b_5 - B_4]^+$	824.434	824.3805	-0.0535
$[b_4 - B_4]^+$	753.3969	753.3832	-0.0137
y_6^+	674.3832	674.3556	-0.0276
$[b_3 - B_4]^+$	654.328	654.3269	-0.0011
y_5^+	575.3148	575.3475	0.0327
$[z_5 - NH_3]^+$	541.2617	541.2237	-0.0380
y_4^+	504.2776	504.2357	-0.0419
$[K^*V - B_4]^+$	452.2867	452.3055	0.0188
y_3^+	417.2456	417.2095	-0.0361
$C_{17}H_{26}N_3O_4^+$	336.1918	336.2023	0.0105
y_2^+	304.1615	304.1637	0.0022
z_2^+	287.135	287.1116	-0.0234
B_1^+	260.1492	260.1505	0.0013
$[B_1 - H_2O]^+$	242.1387	242.1599	0.0212
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1627	0.0390
$[C_1 - C_3H_7O]^+$	202.1074	202.1093	0.0019
$C_7H_{13}O_2^+$	129.091	129.1094	0.0184
$[b_1 - H_2]^+$	112.0757	112.0950	0.0193

Table C.5. Tryptic peptides identified in the two serum albumin protein isoforms by LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate.

Precursor <i>m/z</i> (charge)	Mr(expt)	Mr(calc)	Deviation Da	Missed cleavage	Serum albumin <i>Bos taurus</i> (gi 74267962)	Serum albumin precursor <i>Bos taurus</i> (gi 1351907)	Unique	Peptide
507.8123 (+2)	1013.6101	1013.6121	-0.0020	0	√	√		QTALVELLK
571.8600 (+2)	1141.7055	1141.707	-0.0015	1	√	√		KQTALVELLK
625.3159 (+2)	1248.6172	1248.6139	0.0033	1	√	√		FKDLGEEHFK
642.3654 (+2)	1282.7162	1282.7034	0.0128	0	√	√		HPEYAVSVLLR
653.3464 (+2)	1304.6783	1304.7088	-0.0305	0	√	√		HLVDEPQNLIK
671.3491 (+2)	1340.6836	1340.6799	0.0037	0	√	-	U	TVMENFVAFVVGK
700.3522 (+2)	1398.6899	1398.6853	0.0045	0	-	√		TVMENFVAFVDK
708.3502 (+2)	1414.6859	1414.6803	0.0057	0	-	√		TVMENFVAFVDK + Oxidation (M)
710.3493 (+2)	1418.6841	1418.6864	-0.0023	0	√	√		SLHTLFGDELCK
710.7665 (+2)	1419.5184	1419.5105	0.008	0	√	-	U	ETYGDMADCCAK
720.4098 (+2)	1438.805	1438.8045	0.0006	1	√	√		RHPEYAVSVLLR
722.3251 (+2)	1442.6356	1442.6347	0.0009	0	√	√		YICDNQDTISSK
739.7664 (+2)	1477.5182	1477.516	0.0023	0	-	√		ETYGDMADCCCK
756.4284 (+2)	1510.8422	1510.8355	0.0066	0	√	√		VPQVSTPTLVEVSR
769.4033 (+2)	1536.7921	1536.7936	-0.0016	0	√	-	U	LGEYGFQNELIVR
516.3047 (+2)	1545.8924	1545.8878	0.0046	1	√	√		LKHLVDEPQNLIK
777.8271 (+2)	1553.6396	1553.6457	-0.0061	0	√	√		DDPHACYSTVFDK
784.377 (+2)	1566.7395	1566.7354	0.0041	0	√	√		DAFLGSFLYEYSR
788.8923 (+2)	1575.7701	1575.7603	0.0098	0	√	√		LKPDPTLCLDEFK
820.4754 (+2)	1638.9363	1638.9305	0.0058	1	√	√		KVPQVSTPTLVEVSR
847.9184 (+2)	1693.8223	1693.8167	0.0055	0	√	-	U	MPCAEDYLSLILNR
862.9237 (+2)	1723.8328	1723.8273	0.0055	0	-	√		MPCTEDYLSLILNR

870.9194 (+2)	1739.8243	1739.8222	0.0021	0	-	√		MPCTEDYLSLILNR + Oxidation (M)
874.3575 (+2)	1746.7004	1746.6978	0.0027	0	√	√		YNGVFQECCQAEDK
583.8934 (+2)	1748.6585	1748.6553	0.0033	0	√	√		ECCHGDLLECADDR
599.2800 (+2)	1794.8182	1794.8247	-0.0065	1	√	√		DDPHACYSTVFDKLLK
940.9644 (+2)	1879.9142	1879.9138	0.0004	0	√	√		RPCFSALTPDETYVPK
944.9757 (+2)	1887.9368	1887.9195	0.0173	0	√	√		HPYFYAPELLYYANK
634.6273 (+3)	1900.8601	1900.8625	-0.0025	1	√	√		NECFLSHKDDSPDLPK
636.6473 (+3)	1906.92	1906.9135	0.0065	0	√	√		LFTFHADICTLPDTEK
643.2729 (+3)	1926.797	1926.791	0.006	1	√	√		CCAADDKEACFAVEGPK
673.9933 (+3)	2018.9582	2018.9619	-0.0037	1	√	√		LKPDPNTLCDEFKADEK
682.3489 (+3)	2044.0249	2044.0206	0.0042	1	√	√		RHPYFYAPELLYYANK
705.2984 (+3)	2112.8733	2112.8775	-0.0042	1	√	√		VHKECCHGDLLECADDR
706.6155 (+3)	2116.8246	2116.8136	0.011	1	-	√	U	ETYGDMADCCKEQEPER
749.9835 (+3)	2246.9288	2246.9354	-0.0067	1	√	√		ECCHGDLLECADDRADLAK
820.0702 (+3)	2457.1887	2457.1733	0.0153	1	√	√		AIPENLPPLTADFAEDKDVCK
831.4255 (+3)	2491.2546	2491.257	-0.0024	0	√	√		GLVLIAFSQYLQCCPFDEHVK
1013.4318 (+3)	3037.2735	3037.2416	0.0319	1	√	√		YEATLEECCAKDDPHACYST VFDK

Table C.6. LC-QqTOF-MS/MS analysis of the glycated peptide CCTK*PESER (Lys 463) at m/z 1058.4652 (+2).

Product ion	Calculated m/z	Observed m/z	Deviation (Da)
Y_3^+	1856.7776	1856.7628	-0.0148
Y_2^+	1710.7197	1710.7332	0.0135
Y_1^+	1564.6618	1564.6742	0.0124
Y_0^+	1418.6039	1418.6321	0.0282
$[y_7 - B_4]^+$	1098.5426	1098.5412	-0.0014
$[y_6 - B_4]^+$	997.4949	997.4909	-0.0040
$[b_4 - B_4]^+$	802.3228	802.3246	0.0018
y_5^+	617.2889	617.2909	0.0020
y_4^+	520.2362	520.2245	-0.0117
b_3^+	422.1168	422.1208	0.0040
B_2^+	406.2077	406.2071	-0.0006
y_3^+	391.1936	391.1922	-0.0014
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1908	-0.0010
b_2^+	321.0681	321.0646	-0.0035
y_2^+	304.1615	304.1593	-0.0022
C_1^+	276.1442	276.1435	-0.0007
B_1^+	260.1492	260.1460	-0.0032
$[B_1 - H_2]^+$	258.1336	258.1313	-0.0023
$[B_1 - H_2O]^+$	242.1387	242.1354	-0.0033
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1184	-0.0053
$[C_1 - C_3H_7O]^+$	202.1074	202.1034	-0.0040
$^{24}A_1^+$	184.0968	184.0928	-0.0040
y_1^+	175.1190	175.1144	-0.0046
$[B_1 - C_3H_5O_2]^+$	160.0968	160.0945	-0.0023

Table C.7. Peptides identified in the two serum albumin protein isoforms by LC-ESI-QqTOF-MS/MS analysis of the hapten-BSA glycoconjugate digested with the endoproteinase GluC V8.

Precursor <i>m/z</i> (charge)	Mr(expt)	Mr(calcd)	Deviation Da	Missed cleavage	Serum albumin <i>Bos taurus</i> (gi 74267962)	Serum albumin precursor <i>Bos taurus</i> (gi 1351907)	Unique	Peptide
489.7625 (+2)	977.5104	977.4971	0.0134	0	√	√		KLFTFHAD
535.2620 (+2)	1068.5095	1068.5128	-0.0033	1	√	√		TYVPKAFDE
573.7933 (+2)	1145.5720	1145.5717	0.0003	1	√	√		KSLHTLFGDE
611.2009 (+2)	1220.3872	1220.3784	0.0088	2	-	√		TYGDMADCCCE
622.3434 (+2)	1242.6723	1242.6642	0.0081	1	√	√		DKGACLLPKIE
642.8832 (+2)	1283.7519	1283.7449	0.007	0	√	√		GPKLVVSTQTALA
680.8517 (+2)	1359.6888	1359.6711	0.0177	1	√	√		AKDAFLGSFLYE
681.3677 (+2)	1360.7208	1360.7179	0.0029	0	√	√		KKFWGKLYLE
760.3999 (+2)	1518.7853	1518.7731	0.0122	0	√	√		IARRHPYFYAPE
773.3853 (+2)	1544.7561	1544.7358	0.0202	3	√	√		NLPPLTADFAEDKD
537.6603 (+2)	1609.9591	1609.9403	0.0188	1	√	√		VTKLVTDLTQVHKE
861.4378 (+2)	1720.8611	1720.8460	0.0151	0	√	√		LLYYANKYNGVFQE

610.9461 (+3)	1829.8163	1829.8077	0.0087	2	√	-	U	NFVAFVGKCCAADDKE
931.4739 (+3)	1860.9332	1860.9152	0.018	1	√	√		SLVNRRPCFSALTPDE
630.2817 (+3)	1887.8233	1887.8131	0.0102	3	-	√		NFVAFVDKCCAADDKE
954.4680 (+3)	1906.9214	1906.9135	0.0079	2	√	√		KLFTFHADICTLPDTE
973.3785 (+3)	1944.7424	1944.7441	-0.0017	1	√	√		CCAKDDPHACYSTVFD
709.3707 (+3)	2125.0901	2125.0725	0.0177	2	√	√		LAKYICDNQDTISSKLK E
725.3025 (+3)	2172.8856	2172.8510	0.0346	3	√	-	U	TYGDMADCCAQEPER NE
730.6347 (+3)	2188.8822	2188.8459	0.0362	3	√	-	U	TYGDMADCCAQEPER NE + Oxidation (M)
866.1225 (+3)	2595.3457	2595.3115	0.0342	2	√	√		KLKHLVDEPQNLIKQNC DQFE
960.5398 (+3)	2878.5976	2878.5705	0.0271	0	-	√	U	YGFQNALIVRYTRKVPQ VSTPTLVE
971.4907 (+3)	2911.4504	2911.4174	0.033	3	√	√		SLVNRRPCFSALTPDETY VPKAFDE

Table C.8. LC-QqTOF-MS/MS analysis of the glycated peptide K*VTKCCTE (Lys 495) at m/z 987.9658 (+2).

Product ion	Calculated m/z	Observed m/z	Deviation (Da)
Y_2^+	1569.7023	1569.6618	-0.0405
Y_1^+	1423.6443	1423.6440	-0.0003
Y_0^+	1277.5854	1277.5919	0.0065
$[Y_0 - H_2O]^+$	1259.5764	1259.5710	-0.0054
$[b_7 - B_4]^+$	1130.5338	1130.5222	-0.0116
$[b_5 - B_4]^+$	869.4555	869.4469	-0.0086
y_6^+	798.312	798.3099	-0.0021
$[b_4 - B_4]^+$	709.4243	709.4351	0.0108
y_5^+	697.2644	697.2727	0.0083
$[b_1 - B_4]^+$	581.3293	581.3174	-0.0119
y_4^+	569.1694	569.1620	-0.0074
$[b_2 - B_4]^+$	480.2817	480.2884	0.0067
y_3^+	409.1388	409.1308	-0.0080
B_2^+	406.2077	406.2084	0.0007
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1892	-0.0026
C_1^+	276.1442	276.1415	-0.0027
B_1^+	260.1492	260.1460	-0.0032
$[B_1 - H_2]^+$	258.1336	258.1307	-0.0029
y_2^+	249.1081	249.1040	-0.0041
$[B_1 - H_2O]^+$	242.1387	242.1380	-0.0007
$[C_1 - C_3H_7O]^+$	202.1074	202.1047	-0.0027
$^{24}A_1^+$	184.0968	184.0948	-0.0020
$[B_1 - C_5H_8O_2]^+$	160.0968	160.0912	-0.0056
y_1^+	148.0604	148.0559	-0.0045

APPENDIX D

Additional Figures and Tables for

**CHAPTER 6: Determination of glycation sites by tandem mass spectrometry in a
synthetic lactose-BSA conjugate, a vaccine model prepared by dialkyl squarate
chemistry**

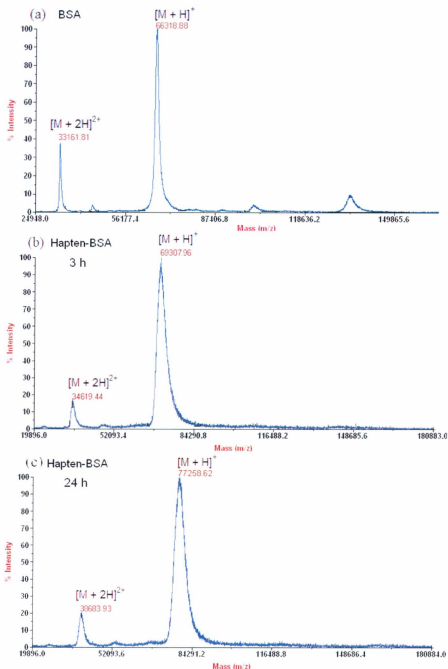


Figure D.1. MALDI-TOF/TOF-MS analysis of the BSA (a) and the lactose-BSA glycoconjugates at the following ratios: 5:1 (b) and 19:1 (c).

Table D.1. LC-ESI-QqTOF-MS/MS analysis of the glycosylated peptide SLGK*VGTR (Lys 455) at m/z 697.3748 (+2).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
Y_1^+	1231.6529	1231.6639	0.0110
Y_0^+	1069.6000	1069.6055	0.0055
$[b_7 - B_2]^+$	895.4889	895.4260	-0.0629
$[y_6 - B_2]^+$	869.4839	869.4918	0.0079
$[y_5 - B_2]^+$	812.4625	812.4715	0.0090
$[b_6 - B_2]^+$	794.4412	794.4752	0.0340
$[b_5 - B_2]^+$	737.3198	737.4277	0.1079
$[b_5 - B_2 - H_2O]^+$	719.4087	719.3841	-0.0246
$[a_5 - B_2]^+$	709.4243	709.4391	0.0148
$[b_4 - B_2]^+$	638.3514	638.3549	0.0035
Y_0^{2+}	535.3037	535.3109	0.0072
$[Y_0 - H_2O]^{2+}$	526.2984	526.3004	0.0020
y_4^+	432.2565	432.2579	0.0014
$[y_6 - B_2 - H_2O]^{2+}$	426.2404	426.2460	0.0056
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1908	-0.0010
y_3^+	333.1881	333.1887	0.0006
y_2^+	276.1666	276.1634	-0.0032
b_3^+	258.1454	258.1401	-0.0053
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1271	0.0034
b_2^+	201.1239	201.1187	-0.0052
y_1^+	175.1190	175.1165	-0.0025
$C_8H_{17}N_2O_2^+$	173.1285	173.1251	-0.0034
B_1^+	163.0606	163.0556	-0.0050
$[B_1 - H_2O]^+$	145.0495	145.0441	-0.0054
$[B_1 - 2H_2O]^+$	127.0390	127.0377	-0.0013

Table D.2. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide VSRSLGK*VGTRCCTKPE (Lys 455) at m/z 837.7418 (+3).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
Y_1^{2+}	1175.0802	1175.0537	-0.0265
Y_0^{2+}	1094.0538	1094.0196	-0.0342
y_8^+	1051.4654	1051.6035	0.1381
$[b_{13} - B_2]^{2+}$	857.4301	857.3536	-0.0765
y_6^+	794.3166	794.3500	0.0334
Y_{13}^+	783.7226	783.7090	-0.0136
Y_0^{3+}	729.7050	729.6955	-0.0095
$[b_{16} - B_2]^{3+}$	680.6874	680.9973	0.3099
$[b_{15} - B_2]^{3+}$	648.3365	648.3185	-0.0180
y_5^+	634.2859	634.2903	0.0044
b_6^+	600.3469	600.3686	0.0217
b_5^+	543.3255	543.3232	-0.0023
b_4^+	430.2414	430.2353	-0.0061
y_3^+	373.2076	373.2002	-0.0074
b_3^+	343.2094	343.2020	-0.0074
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1871	-0.0047
y_2^+	245.1126	245.1056	-0.0070
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1181	-0.0056
b_2^+	187.1083	187.1043	-0.0040
B_1^+	163.0606	163.0527	-0.0079
y_1^+	148.0599	148.0601	0.0002
$[B_1 - H_2O]^+$	145.0495	145.0423	-0.0072

Table D.3. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide VSRSLGK*VGTRCCTK*PE (Lys 455, Lys 463) at m/z 1029.8189 (+3).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
$[y_{10} - B_2]^+$	1459.6657	1459.6788	0.0131
$[y_9 - B_2]^+$	1360.5973	1360.5324	-0.0649
$[y_8 - B_2]^+$	1303.5758	1303.5839	0.0081
$[y_7 - B_2]^+$	1202.5281	1202.4761	-0.0520
$[b_{15} - 2B_2]^{2+}$	1098.0566	1098.0119	-0.0447
$[b_8 - B_2]^+$	1079.6208	1079.6271	0.0063
$[b_7 - B_2]^+$	980.5524	980.5238	-0.0286
$[b_{14} - B_2]^{2+}$	907.9539	907.9193	-0.0346
$[b_{13} - B_2]^{2+}$	857.4301	857.3852	-0.0448
$[b_{12} - B_2]^{2+}$	777.4148	777.3857	-0.0290
$[y_4 - B_2]^+$	726.3657	726.3182	-0.0475
$[b_{11} - B_2]^{2+}$	697.3994	697.3887	-0.0107
$[y_3 - B_2]^+$	625.3180	625.3054	-0.0126
b_6^+	600.3469	600.3208	-0.0261
$[b_9 - B_2]^{2+}$	568.8250	568.7979	-0.0271
b_5^+	543.3255	543.3021	-0.0234
$[b_7 - B_2]^{2+}$	490.7801	490.7766	-0.0034
b_3^+	343.2094	343.2091	-0.0003
$C_{17}H_{26}N_3O_4^+$	336.1918	336.1852	-0.0066
y_2^+	245.1126	245.1060	-0.0066
$C_{11}H_{16}N_3O_2^+$	222.1237	222.1174	-0.0063
B_1^+	163.0606	163.0559	-0.0047
y_1^+	148.0599	148.0556	-0.0043

Table D.4. LC-ESI-QqTOF-MS/MS analysis of the glycated peptide DK*GACLLPK*IE at m/z 1198.5604 (+2).

Molecular ion	Calculated m/z	Experimental m/z	Deviation Da
$[b_7 - B_2]^{-}$	1010.4975	1010.4840	-0.0135
$[b_6 - B_2]^{-}$	897.4135	897.4186	0.0051
$[y_5 - B_2]^{-}$	851.4862	851.4950	0.0088
$[b_5 - B_2]^{-}$	784.3294	784.3192	-0.0102
$[y_4 - B_2]^{-}$	738.4021	738.3985	-0.0036
$[y_4 - C_2]^{-}$	720.3913	720.3922	0.0009
$[y_3 - B_2]^{-}$	641.3494	641.3512	0.0018
$[b_4 - B_2]^{-}$	624.2988	624.2894	-0.0094
$[PK^*I - B_2]^{-}$	591.3495	591.3327	-0.0168
$[PK^*I - B_2 - CO]^{-}$	563.3551	563.3495	-0.0056
$[b_3 - B_2]^{-}$	553.2617	553.2575	-0.0042
$[b_3 - C_2]^{-}$	535.2509	535.2567	0.0058
$[K^*GA - B_2]^{-}$	509.2713	509.2603	-0.0110
$[b_2 - B_2]^{-}$	496.2402	496.2263	-0.0139
$[K^*I - B_2]^{-}$	494.2968	494.2911	-0.0057
$[PK^* - B_2]^{-}$	478.2655	478.2529	-0.0126
$C_{17}H_{26}N_3O_4^{+}$	336.1918	336.1881	-0.0037
y_2^{+}	261.1439	261.1394	-0.0045
$C_{11}H_{16}N_3O_2^{+}$	222.1237	222.1187	-0.0050

